

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/55456>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

EXPRESSION OF THE NUCLEOPROTEIN AND PHOSPHOPROTEIN
GENES OF PNEUMONIA VIRUS OF MICE AND SPECIFIC
INTERACTIONS OF THE GENE PRODUCTS

A DISSERTATION SUBMITTED TO

THE UNIVERSITY OF WARWICK

IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY

JOHN NICHOLAS BARR

NOVEMBER, 1993

TABLE OF CONTENTS

	Page
LIST OF ILLUSTRATIONS	viii
LIST OF TABLES	x
ACKNOWLEDGEMENTS	xi
ABBREVIATIONS	xii
DECLARATION.....	xiv
CHAPTER 1 INTRODUCTION.....	1
1.1: Taxonomic status of the Family <i>Paramyxoviridae</i>	2
1.1.1: Initial classification.	2
1.1.2: Current classification.	4
1.2: The incidence of PVM.	8
1.2.1: Original isolation.	8
1.2.2: Incidence in mammals.....	9
1.2.3: Incidence in humans.....	10
1.3: The structure of the pneumovirus virion.....	10
1.4: The polypeptides of PVM.	11
1.4.1: Introduction	11
1.4.2: The nucleocapsid protein.....	13
1.4.3: The phosphoprotein.....	14
1.4.3.1: The P protein of <i>Paramyxoviridae</i> family members.....	16
1.4.4: The Large protein.....	19
1.4.5: The Matrix protein.	21
1.4.6: The 22K (M2) protein.	24
1.4.7: The 1B and 1C proteins.....	25
1.4.8: The fusion protein.....	27
1.4.9: The attachment protein.....	31
1.4.10: The SH (1A) protein.	37
1.5: Pneumovirus molecular biology.....	39
1.5.1: The pneumovirus genome.	39
1.5.2: The leader and trailer regions.	40
1.5.3: Gene start and end sequences.	41

1.5.4: Intergenic regions.....	43
1.5.5: Pneumovirus RNA synthesis.....	43
1.6: RNA synthesis in negative sense, single stranded RNA viruses.....	45
1.6.1: Introduction.	45
1.6.2: RNA synthesis in VSV.....	45
1.6.3: Components of the VSV nucleocapsid.	46
1.6.4: The model of RNA synthesis in negative sense, single stranded RNA viruses.	49
1.6.5: The modified model of Paramyxoviridae RNA synthesis.	52
1.7: The aims of this study.	55
CHAPTER 2 MATERIALS AND METHODS.....	57
2.1: Growth of PVM in cell culture.....	58
2.1.1: Growth of BSC1 cells.	58
2.1.2: Growth of virus stocks.	58
2.1.3: Plaque assay titration of PVM stocks.	59
2.1.4: Preparation of PVM-infected BSC1 cell material.	59
2.2: Protein analysis.....	60
2.2.1: Polyacrylamide gel electrophoresis.	60
2.2.2: Transfer of proteins from polyacrylamide gel to PVDF membrane.....	61
2.2.3: Western blot analysis.	62
2.2.4: The west-western blot assay.....	63
2.2.5: Densitometric analysis.	64
2.3: Manipulations in bacteria.	64
2.3.1: Bacterial Strains.	64
2.4: Routine manipulations of DNA.....	65
2.4.1: Phenol and chloroform extraction.	65
2.4.2: DNA and RNA precipitations.....	66
2.5: Bacterial transformation and preparation of plasmid DNA.	66
2.5.1: Preparation of competent bacteria.	66
2.5.2: Transformation of <i>E. coli</i> TG2 with plasmid DNA.....	67
2.5.3: M13 DNA transformation.	67
2.5.4: Small scale plasmid preparations.....	68
2.5.5: Large scale plasmid preparations.....	68
2.5.6: Single strand template preparation.	70

2.6: Modification of DNA.....	70
2.6.1: Restriction enzyme digestion of DNA.....	70
2.6.2: End-filling of DNA fragments.....	71
2.6.3: Ligations.....	71
2.7: Electrophoresis of DNA.....	72
2.7.1: Agarose gel electrophoresis.....	72
2.7.2: Polyacrylamide gel electrophoresis.	72
2.7.3: Fragment preparations.....	73
2.7.4: Polymerase chain reaction.....	74
2.8: Characterisation of PVM gene 4 cDNA clones.....	75
2.9: Nucleotide sequence analysis.	76
2.10: Identification of anti-PVM P protein specific monoclonal antibodies.	76
2.11: Expression of radiolabelled polypeptides <i>in vitro</i>	77
2.11.1: Transcription of mRNA <i>in vitro</i>	77
2.11.2: <i>In vitro</i> translation using rabbit reticulocyte lysate.	78
2.11.3: <i>In vitro</i> translation using wheat germ lysate.....	79
2.11.4: Immunoprecipitation of ³⁵ S methionine labelled polypeptides.....	80
2.11.5: Trichloroacetic acid precipitations.....	81
2.12: Expression of polypeptides in bacterial systems.	82
2.12.1: Expression and purification of polypeptides using the Qiagen Express™ system.....	82
2.12.2: Expression of polypeptides using <i>E. coli</i> strain BL21.....	83
2.13: Generation of antisera in laboratory animals.	84
2.13.1: Generation of antisera against the 24 kDa N protein related polypeptide.	84
2.13.2: Generation of antisera against the PVM P gene second ORF protein product.....	84
CHAPTER 3 Determination of the nucleotide sequence of the PVM P gene.....	87
3.1: Construction of M13 vectors for sequencing the PVM P gene.....	88
3.2: The nucleotide sequence of the PVM P gene.....	88
3.3: The predicted products of the PVM P gene.	91
3.4: Discussion.....	95
3.4.1: The PVM P gene.....	95
3.4.2: Comparison of the PVM P protein with the P proteins of other <i>Paramyxoviridae</i> family members.....	95

3.4.3: Comparison of the PVM P protein with other pneumovirus P proteins...	97
3.4.4: The PVM P gene second ORF.....	102
3.4.5: Comparison of the PVM P protein sequence with a protein sequence database.	103
3.4.6: Comparison of the PVM P2 protein with the P gene products of other <i>Paramyxoviridae</i> family members.....	105
CHAPTER 4 Characterisation of the <i>in vitro</i> expression products of the PVM P gene.	106
4.1: Construction of T7 transcription plasmid pGEMP.....	107
4.2: <i>In vitro</i> expression of PVM P protein and P protein related species.....	109
4.2.1: <i>In vitro</i> expression of the PVM P protein.	109
4.2.2: Construction of pGEMP derived transcription plasmids.	109
4.2.3: <i>In vitro</i> expression of P protein related species.....	111
4.3: Discussion.....	114
4.3.1: <i>In vitro</i> expression of the PVM P protein.	114
4.3.2: <i>In vitro</i> expression of other P protein species.	115
4.3.3: Leaky scanning and the expression of the P protein species.....	116
4.3.4: mRNA breakage and expression of the P protein species.....	119
4.3.5: Proteolytic cleavage and expression of P protein species.....	121
CHAPTER 5 Bacterial expression of the PVM P protein and PVM P protein related polypeptides.....	123
5.1: The <i>E. coli</i> BL21 expression system.	124
5.2: <i>E. coli</i> BL21 expression of the P protein and P protein related species.....	124
5.3: Discussion.....	126
5.3.1: The P protein species expressed by <i>E. coli</i> BL21.....	126
5.3.2: Leaky scanning and <i>E. coli</i> BL21 expression.	127
CHAPTER 6 Identification of monoclonal antibodies specific for the PVM P gene....	129
6.1: Introduction.	130
6.2: Monoclonal antibody screening procedure.	131
6.3: Epitope mapping of PVM P protein specific monoclonal antibodies.	131
CHAPTER 7 The PVM P protein species expressed in PVM-infected BSC1 cells.	137
7.1: Western blot analysis of PVM-infected BSC1 cell material.....	138
7.2: Comparison of P protein species expressed in PVM-infected BSC1 cells with P protein species expressed in <i>E. coli</i> BL21.	138
7.3: Discussion.....	141

7.3.1: The 26 kDa and 23 kDa P protein species expressed in PVM-infected BSC1 cells.	141
7.3.2: Other P protein related species expressed in PVM-infected BSC1 cells.	143
CHAPTER 8 Identification of the PVM P gene second ORF protein product (P2) in PVM-infected BSC1 cells.	146
8.1: Expression of P2 in a wheat germ lysate system.....	147
8.2: Translational initiation of the P2 polypeptide <i>in vitro</i>	147
8.3: The generation of antiserum against the P gene second ORF protein product.	150
8.3.1: The Qiagen Express system.....	150
8.3.2: Construction of plasmid pQE32.P2.	151
8.4: Immunological activity of anti-P2 antiserum.....	153
8.5: Identification of P2 in PVM-infected BSC1 cells.	155
8.6: Discussion.....	155
CHAPTER 9 The polypeptide expression products of the PVM N gene.....	159
9.1: PVM N gene expression <i>in vitro</i>	160
9.1.1: Construction of transcription plasmids pGEMN and pGEMNTR.	160
9.1.2: <i>In vitro</i> translation products of pGEMN and pGEMNTR.	160
9.2: Generation of antiserum against the 24 kDa PVM N protein species.	164
9.2.1: Construction of pQE32.24.....	164
9.2.2: Immunisation of rabbit with purified 24 kDa N protein species.....	164
9.2.3: Identification of the 24 kDa species in PVM-infected BSC1 cell material.....	168
9.3: Discussion.....	170
9.3.1: Expression products of the PVM N gene.	170
9.3.2: Expression of the 24 kDa N protein species.....	170
9.3.3: The 30 kDa N protein species expressed in PVM-infected BSC1 cells.	175
CHAPTER 10 Interaction between the N and P proteins of PVM.	178
10.1: The west western binding assay	179
10.2: The full-length PVM N protein is able to interact with the PVM P ₃₉ protein <i>in vitro</i>	180
10.3: Identification of domains on the PVM N protein responsible for binding to renatured P protein.....	180
10.3.1: Introduction.	180
10.3.2: Transcription plasmids constructed using N gene cDNA.	183
10.3.3: <i>In vitro</i> expression of N protein related polypeptides.....	189

10.3.4: The ability of the altered N protein species to bind to PVM P protein.192

10.4: Identification of domains on the PVM P protein that are involved in binding to PVM N protein.197

10.4.1: Introduction.197

10.4.2: Construction of plasmids to express carboxyl-terminated P protein species.197

10.4.3: The abilities of the altered P protein species to bind to PVM N protein.199

10.5: Discussion.....202

10.5.1: PVM N protein binds to immobilised PVM P protein.....202

10.5.2: The regions of the N protein that are involved in binding P protein...202

10.5.3: Regions of the P protein that are involved in binding N protein.....204

CHAPTER 11 Concluding remarks.208

11.1: Introduction.209

11.2 The expression characteristics of the PVM P gene.....209

11.3: The expression characteristics of the PVM N gene.....212

11.4: Possible roles for the additional N and P gene encoded proteins.....212

CHAPTER 12 BIBLIOGRAPHY215

LIST OF ILLUSTRATIONS

Figure	Page
Figure 1. The gene order in the three <i>Paramyxoviridae</i> family genera.	5
Figure 2. Diagrammatical representation of the PVM virion.	12
Figure 3. P gene expression in SEN.	18
Figure 4. Structural features of the F protein that are conserved amongst <i>Paramyxoviridae</i> family members.	29
Figure 5. Elution profile of MBS-activated keyhole limpet haemocyanin.....	86
Figure 6. Region of the PVM P gene represented by M13 subclones constructed to determine the P gene nucleotide sequence.....	89
Figure 7. Nucleotide sequence of the PVM P gene.....	90
Figure 8. Alignment between the amino acid sequences of the HRSV and PVM P proteins.....	98
Figure 9. Matrix plot of identities between PVM P protein and HRSV P protein.	99
Figure 10. Hydropathy comparison between the P proteins of HRSV and PVM.....	101
Figure 11. Construction of plasmid pGEMP.	108
Figure 12. <i>In vitro</i> expression of P gene mRNA.....	110
Figure 13. Construction of T7 transcription plasmids pGEMP37, pGEMP26 and pGEMP23.....	112
Figure 14. Mapping of AUG codons responsible for initiation of <i>in vitro</i> synthesised P protein species	113
Figure 15. Nucleotide context of the first five AUG codons of the PVM P gene P protein ORF.....	118
Figure 16. Expression of P protein and P protein related polypeptides in <i>E. coli</i> BL21.....	125
Figure 17. Screening of anti-P protein monoclonal antibodies.....	132
Figure 18. Mapping the regions where anti-P protein monoclonal antibodies bind.	135
Figure 19. Characterisation of anti-P protein monoclonal antibodies.....	136
Figure 20. P protein related polypeptides expressed in PVM infected cells	139
Figure 21. The PVM P gene expresses at least three polypeptides <i>in vivo</i>	140
Figure 22. <i>In vitro</i> expression of the P2 polypeptide.	148
Figure 23. Construction of plasmid pQE32.P2.	152
Figure 24. The PVM P gene second ORF encodes a polypeptide <i>in vivo</i>	154
Figure 25. Construction of pGEMN and pGEMNTR.	161
Figure 26. <i>In vitro</i> expression products of pGEMN and pGEMNTR	163

Figure 27. Construction of pQE23.24.	165
Figure 28. Purification of the truncated (24 kDa) N protein polypeptide.	166
Figure 29. Recognition of the bacterially expressed truncated (24 kDa) N protein species using the anti-24 kDa N protein antiserum.	167
Figure 30. The N protein related species in PVM-infected BSC1 cells.....	169
Figure 31. The nucleotide context of the first 18 AUG codons of the PVM N gene.....	171
Figure 32. SQUIGGLES plot of PVM N gene mRNA	173
Figure 33. Full-length N protein binds to full-length P protein.....	181
Figure 34. Construction of T7 transcription plasmids using PCR to express altered N proteins for use in the west western assay.	182
Figure 35. Construction of plasmid pGEMN*	184
Figure 36. Construction of transcription plasmids with internally deleted N protein coding regions.	187
Figure 37. Diagrammatical representation of N gene cDNA fragments used to express N proteins for use in the west western assay.....	188
Figure 38. Expression of altered N proteins for use in the west western assay (I).....	190
Figure 39. Expression of altered N proteins for use in the west western assay (II).....	191
Figure 40. Determination of the regions of the PVM N protein that are involved in binding to the PVM P protein (I).	193
Figure 41. Determination of the regions of the PVM N protein that are involved in binding to the PVM P protein (II).	194
Figure 42. Construction of pGEMPSty, pGEMPStu and pGEMPHinc.	197
Figure 43. The regions of the P protein represented by P gene constructs for use in the west western assay.	198
Figure 44. Western blot of altered P proteins for use in the west western assay.....	200
Figure 45. Determination of regions of the PVM P protein that are involved in binding to the PVM N protein.....	201

LIST OF TABLES

Table	Page
Table 1: Members of the <i>Paramyxoviridae</i> family..	3
Table 2: Gene start and end sequences for the 10 genes of HRSV and PVM.	42
Table 3. The amino acid composition of the PVM P protein.	92
Table 4. The amino acid composition of the PVM P2 protein.	94

ACKNOWLEDGEMENTS

First and foremost I would like to acknowledge the vast amounts of patience, guidance and support Andy Easton has given me throughout my PhD study period. Next, I would like to extend gratitude toward my learned colleagues Pat Cane and Phil Chambers for much advice on technical matters. I would also like to thank my not-so-learned (Jo and Kate), ridiculous (Dave P.) and mentally deformed (Jaspal 'waz' Randhawa, Kurt 'spurt' Tolley and David 'muffy' Mathews) lab colleagues for their spiritually uplifting presence. NOT! Thanks also to the MRC, the DSS, my parents and my brother for financial support.

ABBREVIATIONS

BRSV	bovine respiratory syncytial virus
BPIV	bovine parainfluenza virus
cpm	counts per minute
Ci	Curie
Da	Dalton
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetra acetic acid
<i>E. coli</i>	<i>Escherichia coli</i>
GMEM	Glasgow minimal essential medium
g	gravity
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethane-sulphonic acid
HIV	human immunodeficiency virus
HRSV	human respiratory syncytial virus
HPIV	human parainfluenza virus
IPTG	isopropyl- β -D-thiogalactoside
IVT	<i>in vitro</i> translation
kb	kilo base
kDa	kilo daltons
KLH	Keyhole limpet haemocyanin
MBS	m-maleimidobenzoyl N-hydroxysuccinimide
MV	Measles virus
mRNA	messenger ribonucleic acid
MuV	mumps virus
NDV	Newcastle disease virus

OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFU	plaque forming unit
PVDF	polyvinylidene difluoride
PVM	pneumonia virus of mice
RNA	ribonucleic acid
RNase	ribonuclease
SBB	standard binding buffer
SDS	sodium dodecyl sulphate
SEN	Sendai virus
SV5	simian virus 5
SV41	simian virus 41
Tris	Tris-hydroxymethylenediamine
TAE	Tris acetate
TBE	Tris borate
TBS	Tris buffered saline
TCA	trichloroacetic acid
TEMED	N,N,N,N' tetramethylenediamine
TRTV	turkey rhinotracheitis virus
w/v	weight/volume
v/v	volume/weight

DECLARATION

I declare that all the work presented in this thesis, except where specifically stated, was original research performed by myself under the supervision of Dr. Andrew Easton at the Department of Biological Sciences, University of Warwick. None of this work has previously been submitted for any other degree. All sources of information have been acknowledged by means of reference.

ABSTRACT

Following the molecular cloning of the PVM genome, the opportunity to study the individual genes and proteins of PVM has arisen. This study investigated the nucleocapsid (N) gene and the phosphoprotein (P) gene of PVM and attempted to characterise the polypeptide products expressed from the N and P genes both *in vitro* and in PVM-infected cells. The ability of the PVM N and P proteins to interact with each other was also investigated.

The nucleotide sequence of the PVM P gene was determined to be 903 nucleotides in length and shown to comprise a long open reading frame capable of encoding the 295 amino acid long P protein and also a smaller second ORF with the potential to express a polypeptide 137 amino acids in length. The PVM P protein shows overall amino acid homology of 35.3%, 35.6% and 28.3% to the P proteins of pneumovirus members HRSV, BRSV and TRTV respectively. The PVM P gene contrasts with the P genes of other pneumovirus genus members which do not possess any extensive alternative ORFs.

Both the N and P genes of PVM were shown to be capable of directing the synthesis of more than one polypeptide product both *in vitro* and in PVM-infected BSC1 cells. mRNA transcribed from the PVM P gene long ORF directed the *in vitro* expression of the 39 kDa P protein and four additional polypeptides. By constructing transcription plasmids that contained 5' terminally truncated P gene cDNA insets, these polypeptides were determined to be expressed by translational initiation on internal P gene initiation codons. Western blot analysis determined that in addition to the PVM P protein, two of these *in vitro* expressed P protein species, with molecular weights of 26 kDa and 23 kDa, were expressed in PVM-infected BSC1 cells and this observation was supported by the results of anti-P protein monoclonal antibody epitope mapping studies. The ability of the PVM P gene to direct the expression of P protein related polypeptides from internal initiation codons is a feature not yet described for any other pneumovirus member.

By immunising rats with a synthetic peptide, antiserum specific for the second ORF polypeptide product (P2) was generated. Western blot analysis using this anti-P2 antiserum identified a species thought to represent P2 in PVM-infected BSC1 cell material. The ability of the PVM P gene to express a polypeptide from an alternative ORF is a feature common to the P genes of most other morbilliviruses and paramyxoviruses.

mRNA transcribed from PVM N gene cDNA was able to direct the *in vitro* translation of the 43 kDa N protein and also a highly abundant polypeptide with a molecular weight of 24 kDa which was shown to be expressed by way of internal initiation on the fifth N gene AUG codon of the N gene sequence. The 24 kDa N protein related polypeptide was expressed in *E. coli*, purified, and used to immunise a rabbit for the production of anti-24 kDa polypeptide antiserum. Western blot analysis using this antiserum with PVM-infected BSC1 cells detected the 43 kDa N protein, a highly abundant 30 kDa N protein related species, but not the 24 kDa polypeptide. The precise identity of the 30 kDa polypeptide was not determined. Possible mechanisms which could account for the expression of the protein products of the N and P genes are discussed.

By using a protein blotting technique the interaction that occurs between the N and P proteins of PVM was investigated. The P protein binding affinities of *in vitro* expressed truncated N proteins suggested that many regions of the N protein are co-

operatively involved in the binding process, although some regions contributed more than others. The N protein of Sendai virus is believed to bind to the Sendai virus P protein in a similar way. It was also determined that both the amino and the carboxyl-terminal regions of the PVM P protein were found to be essential for binding to N protein. This contrasts with the situation determined for Sendai virus in which 344 P protein amino-terminal amino acids were found to be dispensable for binding N protein.

CHAPTER 1

INTRODUCTION.

1.1: TAXONOMIC STATUS OF THE FAMILY *PARAMYXOVIRIDAE*.

1.1.1: Initial classification.

In 1975, the International Committee on Taxonomy of Viruses (ICTV) suggested that the family *Paramyxoviridae* should subsume three genera, namely paramyxovirus, morbillivirus and pneumovirus. The viruses that are currently classified in these three genera are shown in table 1.

The division of the family into three genera at this time was based upon a consideration of physical details of both nucleocapsid structures and envelope surface projections determined by electron microscope investigations, and also on antigenic and biochemical aspects of the large glycoprotein moiety present in the virion envelope. Members of the morbillivirus and paramyxovirus genera possess nucleocapsids with a diameter of 18 nm and a helical pitch of 5-6 nm whereas the pneumovirus nucleocapsid has been determined to have a diameter of 13.5 nm and a helical pitch of 6.5 nm. The pneumovirus members possess surface projections with a length of 12 nm in contrast to the paramyxovirus and morbillivirus surface projections which have a length of 8 nm. The members of the paramyxovirus genera all possess antigenically cross-reacting glycoprotein epitopes which distinguish them from the pneumovirus and morbillivirus genus members. The paramyxovirus genus members are also distinguishable by the ability of their major glycoprotein to exhibit both haemagglutinin and neuraminidase activity whereas the major glycoprotein of neither the morbillivirus nor the pneumovirus genera members displays neuraminidase activity.

paramyxovirus	morbillivirus	pneumovirus
Human parainfluenza viruses 1, 2, 3, and 4	Measles virus	Human respiratory syncytial virus
Mumps virus	Canine distemper virus	Bovine respiratory syncytial virus
Newcastle disease virus	Phocine distemper virus	Pneumonia virus of mice
Simian virus 5	Rinderpest virus	Turkey rhinotracheitis virus
Simian virus 41	Peste des petits ruminants virus	
Bovine parainfluenza viruses		
Sendai virus (murine PIV-1)		
La-Piedad-Michoacan- Mexico virus		
Avian parainfluenza viruses		

Table 1: Members of the *Paramyxoviridae* family. Adapted from Galinski and Wechsler (1991).

1.1.2: Current classification.

(a) Introduction

Since the initial classification of these genera, techniques have emerged which have allowed analysis of the genome and individual genes of these viruses at the molecular level. The gene order of the three *Paramyxoviridae* family genera are shown in figure 1. The organisation and the nucleotide sequence of the genome of *Paramyxoviridae* family members are more meaningful criteria on which to deduce evolutionary relatedness than the macromolecular observations that were used in the initial classification, described above. Genetic analysis of *Paramyxoviridae* family members has suggested that the family can be separated into four groups, namely the pneumoviruses, the morbilliviruses and two subgroups of the paramyxovirus genus. Analysis of these viruses at the molecular level has emphasised the uniqueness of the pneumoviruses and also confirmed the relatedness that exists between the morbillivirus and paramyxovirus genera.

(b) Pneumovirus characteristics.

The pneumovirus genus presently comprises four viruses namely pneumonia virus of mice (PVM), human respiratory syncytial virus (HRSV), bovine respiratory syncytial virus (BRSV) and turkey rhinotracheitis virus (TRTV). These pneumoviruses possess a markedly different genomic structure to the other members of the *Paramyxoviridae* family, for example their genomes have ten genes compared with six or seven for the paramyxovirus or morbillivirus genomes (figure 1). In addition, the polypeptides which are encoded by the pneumovirus genome are generally smaller than their homologues in other paramyxoviruses and morbilliviruses and the polypeptides display virtually no amino acid similarity to those of other viruses outside the genus

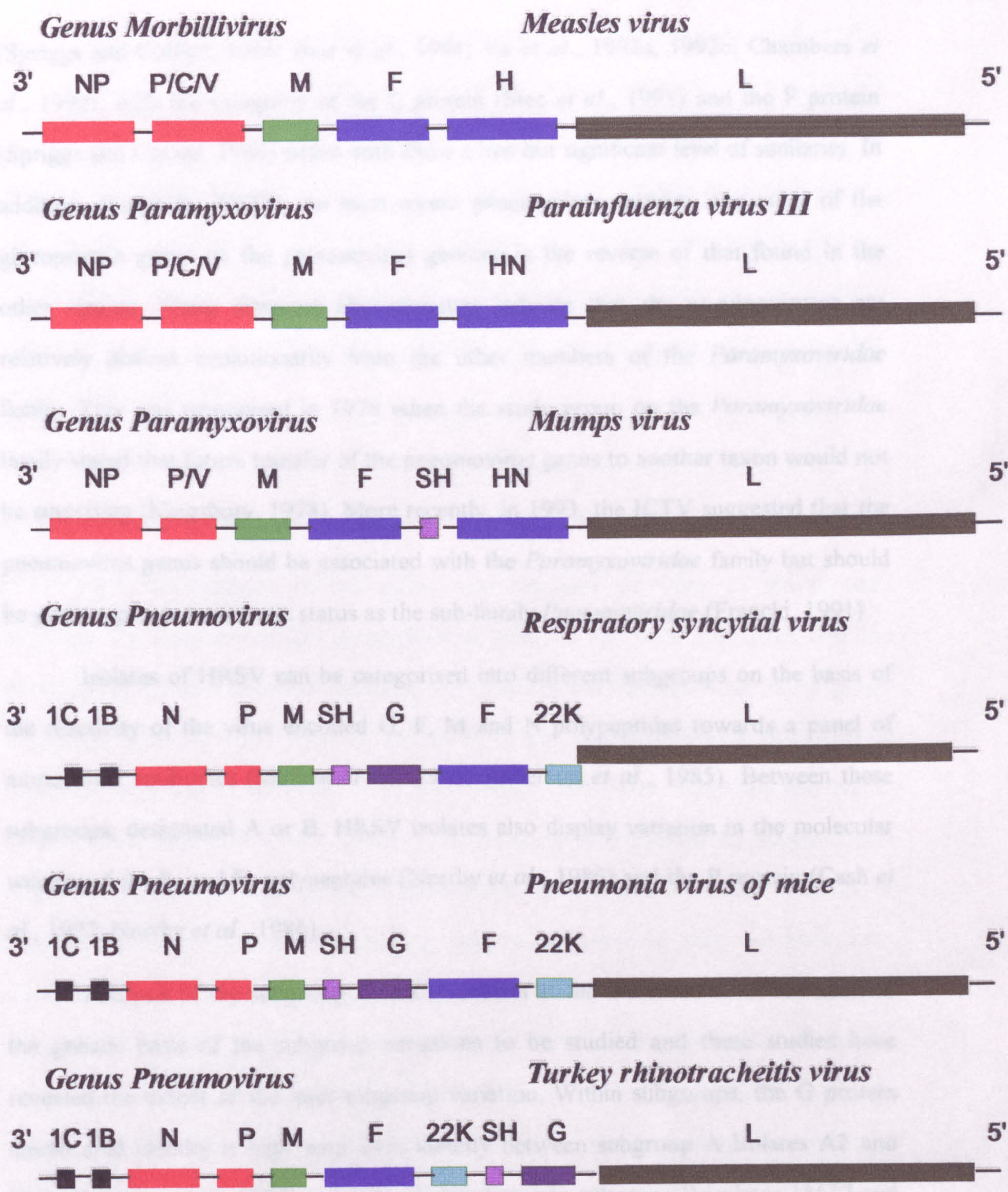


Figure 1. Gene order of representatives of members of the different genera of the family *Paramyxoviridae*

(Spriggs and Collins, 1986; Barr *et al.*, 1991; Yu *et al.*, 1992a, 1992c; Chambers *et al.*, 1992), with the exception of the L protein (Stec *et al.*, 1991) and the F protein (Spriggs and Collins, 1986) which both show a low but significant level of similarity. In addition, except for TRTV, the most recent pneumovirus member, the order of the glycoprotein genes on the pneumovirus genome is the reverse of that found in the other viruses. These disparate characteristics indicate that the pneumoviruses are relatively distinct evolutionarily from the other members of the *Paramyxoviridae* family. This was recognised in 1978 when the study group on the *Paramyxoviridae* family stated that future transfer of the pneumovirus genus to another taxon would not be surprising (Kingsbury, 1978). More recently, in 1993, the ICTV suggested that the pneumovirus genus should be associated with the *Paramyxoviridae* family but should be given separate taxonomic status as the sub-family *Pneumoviridae* (Francki, 1991).

Isolates of HRSV can be categorised into different subgroups on the basis of the reactivity of the virus encoded G, F, M and N polypeptides towards a panel of monoclonal antibodies (Mufson *et al.*, 1985; Anderson *et al.*, 1985). Between these subgroups, designated A or B, HRSV isolates also display variation in the molecular weights of the F₁ and F₂ polypeptides (Norrby *et al.*, 1986) and the P protein (Cash *et al.*, 1977; Norrby *et al.*, 1986).

Analysis of the subgroup A and B isolates at the molecular level has allowed the genetic basis of the subgroup variations to be studied and these studies have revealed the extent of the inter-subgroup variation. Within subgroups, the G protein amino acid identity is high with 94% identity between subgroup A isolates A2 and Long (Johnson *et al.*, 1987) and 98% identity between subgroup B isolates 18537 and 8/60 (Sullender *et al.*, 1990). This contrasts sharply with the level of amino acid identity identified between strains A2 and 18537 which was only 53% (Johnson *et al.*, 1987). Although other HRSV proteins do show difference between subtype isolates, the variation is less than that displayed by the G protein. The SH proteins from

different subgroups show an overall amino acid identity of 76% (Collins *et al.*, 1990) and for the F protein the inter-subgroup identity between two isolates (A2 and 18537) was found to be 91% (Johnson and Collins, 1988). The inter-subgroup amino acid identity (for strains A2 and 18537) of the internal virion components is consistently high (N [96%], 1B [92%], 22K [92%], P [90%], 1C [87%], Reviewed by Collins, 1991). The intra-group amino acid conservation is considerably higher than inter-group conservation for all proteins which suggests that the two subgroups have been evolving for some time.

(c) Division of the paramyxovirus genus.

From comparisons of the amino acid sequence of many of the proteins expressed by members of the paramyxovirus genus it has been concluded that the genus can be divided into two distinct groups, namely the homogeneous PIV-1 group containing HPIV-1, SEN, HPIV-3, and BPIV-3 and the more diverse PIV-2 group which is composed of HPIV-2, SV5, SV41, MuV, HPIV-4A, HPIV-4B and NDV. (Rima, 1989; Miyahara *et al.*, 1992). This grouping is supported by several other observations, such as immunological cross-reactivity and electrophoretic mobility of the member virus P proteins. The strategy used to express the P gene also divides the paramyxoviruses into these two groups. The P protein of the PIV-1 group is encoded by a faithfully copied mRNA and the V protein is translated from a mRNA in which a non-templated G residue has been inserted (section 1.4.3.1). In contrast the P protein of the PIV-2 group is translated from an mRNA in which two non-templated G residues are inserted and the mRNA faithfully copied from the P gene encodes the V protein. Additional reinforcement of this division is seen in the gene start sequences and the intergenic sequences of the paramyxovirus members. In the viruses of PIV-1 group these sequences show a fair degree of conservation, whereas in the PIV-2 group, these sequences are variable.

(d) Paramyxovirus and morbillivirus relationships.

The morbillivirus genus members are almost identical to the members of the PIV-1 group of paramyxoviruses in regard to genetic map, arrangement and sizes of ORFs and sequences of intergenic regions (Collins, 1991). In addition the morbillivirus N, M, F, and L proteins share sequence relatedness to their paramyxovirus counterparts. The degree of relatedness shown by the morbilliviruses to each of these two groups depends on which of the viral proteins are being compared. For example, on the strength of amino acid sequence similarities of the N proteins, the morbillivirus members are more related to the PIV-1 group, whereas if the M protein amino acid sequences are compared, the morbillivirus members are more related to the PIV-2 group. Recent phylogenetic analyses based on the nucleotide sequence of the L protein gene and the P and V protein gene of many paramyxovirus and morbillivirus members have suggested that measles virus (MV) is more closely related to the PIV-1 group of paramyxoviruses than the PIV-1 group is towards NDV, a PIV-2 group member (Stec *et al.*, 1991).

1.2: THE INCIDENCE OF PVM.

1.2.1: Original isolation.

The isolation of a novel virus capable of inducing a fatal pneumonia in mice was reported in 1939 (Horsfall and Hahn, 1939). These workers were attempting to isolate a virus from the pharyngeal washings of patients with various non-influenzal diseases. The washings were inoculated into mice *via* the nasal route and lung suspensions of inoculated mice were serially passed to other mice. To act as a control, lung homogenates of apparently healthy mice were also serially passed to other mice. In both series, passage of lung material resulted in the appearance of pulmonary consolidation in subsequent lung biopsies. In a later study, a total of 31 separate series

were carried out. From 65% of these series an agent responsible for producing transmissible pneumonia was isolated. Death resulted on average after the fifth passage but occurred as early as the third passage and the cultivation of a disease causing agent was believed to be due to the presence of a latent virus in the original animals (Horsfall and Hahn, 1940).

1.2.2: Incidence in mammals.

Early studies found evidence for infection in several other species of animal following intranasal inoculation with PVM. This observation led to the conclusion that PVM was only pathogenic towards mice (Horsfall and Hahn, 1940). Evidence for the occurrence of a virus related to PVM in hamsters (Pearson and Eaton, 1940) and in both cotton rats and rabbits (Eaton and Van Herick, 1944) was reported soon after. Horsfall and Curnen (1946), later demonstrated the presence of PVM neutralising antibody in rabbits, cotton rats, monkeys (*Macaca mulatta*), humans, chimpanzees, hamsters and guinea pigs, but not in chickens or mongooses. Since the presence of such antibodies suggested a previous encounter with PVM, it was postulated that PVM or a related virus was widely distributed amongst mammalian species. More recent studies have identified PVM as being prevalent in laboratory animal colonies of both hamsters and guinea pigs (Carthew *et al.*, 1978), although Gannon and Carthew (1980) were unable to detect the presence of PVM in two guinea pig colonies which were tested. Several studies have found that PVM replication is restricted to the lungs (Carthew and Sparrow, 1980; Weir *et al.*, 1988; Horsfall and Hahn, 1940; Horsfall and Ginsberg, 1951). However a more recent study by Smith *et al.*, (1984) determined that replication can also occur in the upper respiratory tract. PVM may therefore not be strictly pneumotropic.

1.2.3: Incidence in humans.

In initial studies, the presence of PVM neutralising antibody in the serum taken from laboratory workers in contact with mice was reported. Humans who had no such contact with mice also possessed neutralising antibodies (Horsfall and Hahn, 1940). An additional study by Horsfall and Curnen (1946) demonstrated the presence of neutralising antibody in 27% of a large group of humans. Horsfall *et al.*, (1943) also reported the recovery of a virus believed to be PVM from patients with primary atypical pneumonia. Unfortunately these virus isolates are no longer available for clarification of this potentially important finding.

A more recent study performed by Pringle and Eglin (1986) determined that in a random sample of adult sera, 75% possessed PVM neutralising antibody. A study of 108 paired serum samples was undertaken and in 4 instances apparent seroconversion occurred, indicating that PVM, or an antigenically related virus, may contribute to the load of human respiratory disease, although the clinical significance is unclear.

1.3: THE STRUCTURE OF THE PNEUMOVIRUS VIRION.

With the use of halogenated deoxyuridines and acridine orange, PVM was demonstrated to possess a single-stranded RNA genome (Harter and Chopin, 1967). Electron microscopic examination of PVM morphology identified the presence of an outer membrane surrounding both circular and filamentous PVM particles. This membrane was determined to be host cell derived due to the observation that the budding virion was continuous with the cell membrane. The presence of an internal thread-like structure within the filamentous particles was identified, and the diameter of this internal component of the filaments was found to be between 12 and 15 nm. The circular particles were extremely abundant and because of this were thought not to

represent cross-sectioned filaments (Compans *et al.*, 1967). In a comparative study of the morphology of HRSV and PVM, Berthiaume *et al.*, (1974) described the PVM virion in more detail. The internal, thread-like, nucleocapsid component of the filamentous PVM virions was measured with increased resolution and determined to have a diameter of 13.5 nm and a helical pitch of 6.5 nm. The thread-like nucleocapsid structures were absent from the circular virion forms but due to the appearance of dots in their centres, these circular forms were believed to represent filament cross sections.

A diagrammatical representation of the PVM virion is shown in figure 2. The internal nucleocapsid component of the virion is composed of the RNA genome associated with the nucleocapsid protein (N), the phosphoprotein (P) and the large protein (L). Surrounding the internal nucleocapsid is the virion envelope which is the location for the attachment glycoprotein (G), the fusion glycoprotein (F) and possibly the small hydrophobic protein (SH). The matrix protein (M) and the 22K (or M2) protein are believed to be located on the inner face of the outer membrane. The remaining two PVM polypeptides, 1B and 1C, are believed to be non-structural by analogy with HRSV.

1.4: THE POLYPEPTIDES OF PVM.

1.4.1: Introduction

In the instances where little is known about a PVM polypeptide, the corresponding polypeptide of HRSV is discussed in detail. Due to the close relatedness of these viruses, their protein components are expected to be very similar in structure and function. The polypeptides are described in turn beginning with the

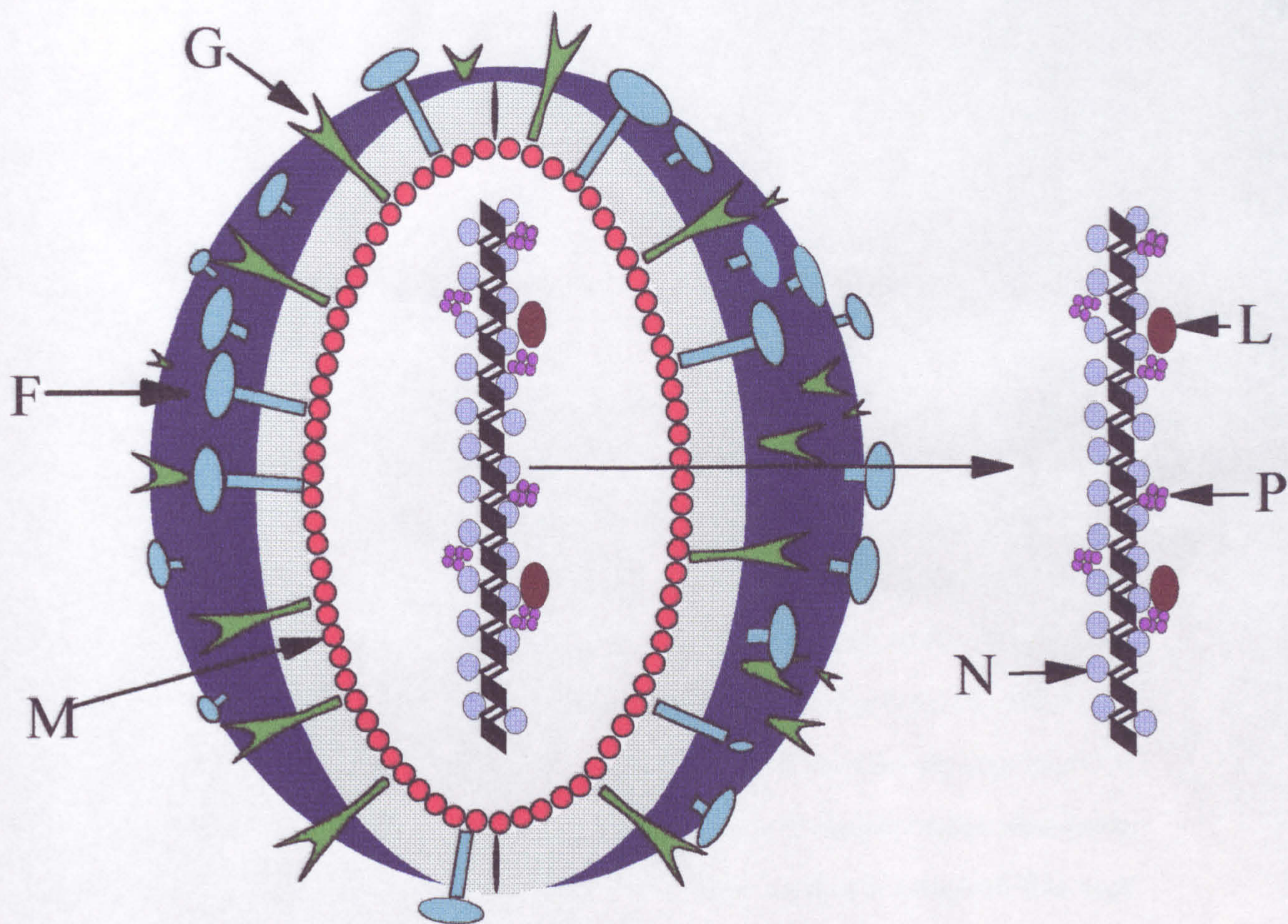


Figure 2. Diagrammatic representation of the PVM virion showing the major protein components.

internal nucleocapsid associated components (N, P and L), followed by the matrix (M) and 22K proteins, then the non-structural 1C and 1B proteins and finally the membrane proteins F, G and SH.

1.4.2: The nucleocapsid protein.

The PVM N gene, which is the third gene on the PVM genome, is 1215 nucleotides in length and has the capacity to encode the 393 amino acid long PVM N protein (Barr *et al.*, 1991). The molecular weight of the N protein predicted from the amino acid sequence (43,141 Da), matches the molecular weight calculated for this protein by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE; 43 kDa). Comparison of the amino acid sequence of the N proteins of PVM and HRSV reveals that they are extremely similar with an overall identity of 60%. Especially high identity between these two proteins is seen for the region containing amino acids 245 to 315 for which 68 out of 71 amino acids are identical. The high degree of conservation that is seen between these proteins is consistent with the observation that the N proteins of PVM and HRSV are serologically related (Gimenez *et al.*, 1984). The PVM N protein also shows a good level of identity with the BRSV N protein (Samal *et al.*, 1991) and the TRTV N protein (45%; A. Easton, pers. comm.) and thus it seems that the pneumovirus N proteins are well conserved. The pneumovirus nucleocapsid proteins are very much shorter than the N proteins of other *Paramyxoviridae* family members, and this feature complicates the search for amino acid and structural homologies that may exist between these proteins. However, limited regions of apparent identity have been identified between the PVM N protein and the N proteins of various other non-segmented, negative sense, single stranded RNA viruses such as the rhabdovirus, vesicular stomatitis virus, and the filovirus, Ebola virus (Barr *et al.*, 1991). Furthermore, the hydropathy profiles of N proteins from PVM and from members of each of the distinct groups of non-segmented,

negative sense, single stranded RNA viruses resemble each other in the same region that displays greatest sequence similarity between the N proteins of the paramyxovirus and morbillivirus genera (Galinski *et al.*, 1986). In accordance with the assumption that the pneumovirus N protein shares functional properties with N proteins of other *Paramyxoviridae* family members, the PVM N protein would be expected to represent the major structural protein of the viral nucleocapsid which is responsible for encapsidation of the RNA genome, a function which requires the ability for the N protein to not only interact with the negative stranded genome molecule (vRNA) but also to other N protein molecules. In addition, the functional N protein must allow the RNA dependant RNA polymerase to access the vRNA and this function supposes that the N protein must be capable of interacting with the polymerase components, namely the L and P proteins. In accordance with this scheme, recent work has shown that the BRSV N protein is capable of interacting with the BRSV P protein (Samal *et al.*, 1993), and work presented in this thesis shows that the PVM N protein is able to interact with the PVM P protein (chapter 10). The N and P proteins of HRSV have also been shown to form an association when co-expressed from cDNA clones using a vaccinia transient expression system (Garcia *et al.*, 1993).

1.4.3: The phosphoprotein.

Determination of the PVM P gene nucleotide sequence is reported in this thesis and for this reason the results of the sequence analysis are not extensively reported in this section, but are fully described in Chapter 3.

The PVM P gene is 903 nucleotides in length, and directs the synthesis of the 295 amino acid PVM P protein from a open reading frame that stretches from an AUG codon at nucleotide position 10, to an UAG codon at nucleotide position 895. The P gene also contains a second open reading frame which begins at an AUG codon at nucleotide position 131 and which ends at an UAG codon located at nucleotide

position 542. This smaller ORF has the capacity to encode a polypeptide 137 amino acids in length. The P protein genes of HRSV and TRTV do not possess a second open reading frame (Satake *et al.*, 1984; Lambden, 1985; A. Easton, pers. comm.). The presence of the second ORF in the P gene of PVM means that possession of a P gene with multiple ORFs is not an exclusive property of the P genes of members of the paramyxovirus and morbillivirus genera. The predicted molecular weight of the PVM P protein is 32,916 Da. This figure is not in close agreement with the observed molecular weight of the P protein as estimated by SDS-PAGE analysis, which is 39 kDa (Cash *et al.*, 1979, Ling and Pringle, 1989a). Aberrant migration in SDS-PAGE seems to be a common feature of the P proteins of *Paramyxoviridae* family members and this phenomenon is believed to stem from the inability of SDS to maximally bind to the P proteins on account of their characteristically negative charge (Caravokyri and Pringle, 1992). *In vivo* labelling studies determined that the PVM P protein is phosphorylated (Ling and Pringle, 1989a). The HRSV P protein is also phosphorylated and a study by Navarro *et al.*, (1991) mapped the sites of the phosphorylated residues to the central region of the molecule. By analogy with the P proteins of other *Paramyxoviridae* family members, the PVM P protein is believed to represent a component of the RNA dependent RNA polymerase. In these viruses, phosphorylation of the P protein is involved in the regulation of transcriptional activity of the viral polymerase complex, and thus phosphorylation of the HRSV and PVM P proteins may perform a similarly crucial function.

In the light of finding that the PVM P gene has a second extensive ORF, it is informative to understand the strategies used in the expression of the multiple ORFs of the P genes of other *Paramyxoviridae* family members.

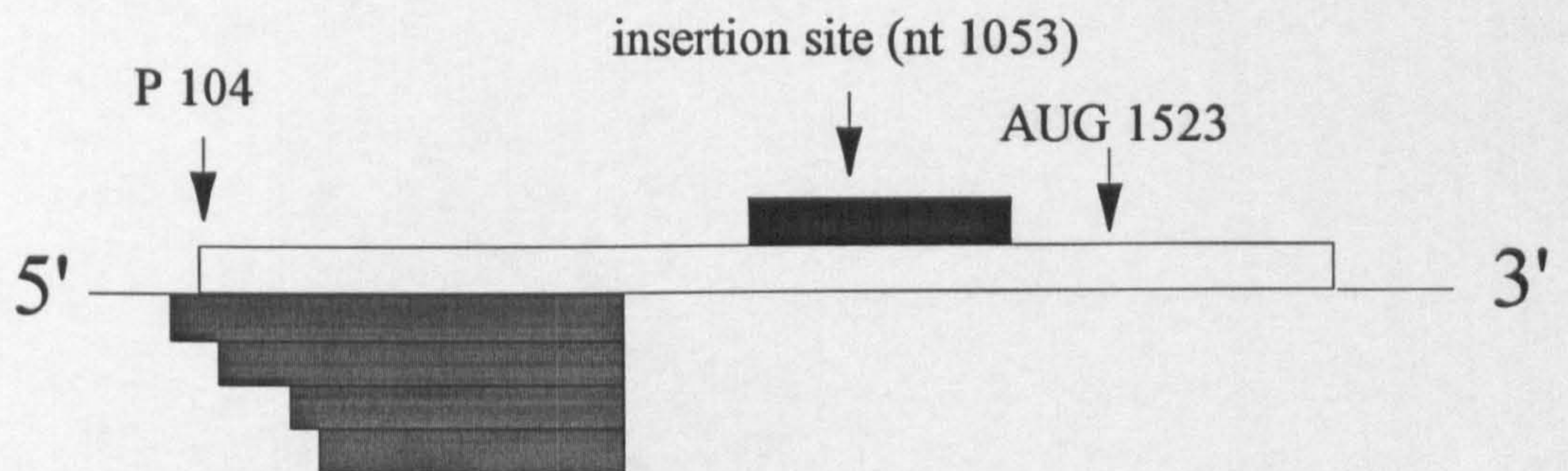
1.4.3.1: THE P PROTEIN OF *PARAMYXOVIRIDAE* FAMILY MEMBERS.

All paramyxovirus and morbillivirus P genes analysed to date are capable of directing the expression of several polypeptide products. In order to do this the P genes are able to utilise reading frames other than that which encodes the major P protein, or are able to switch between ORFs by RNA editing. The mechanism of RNA editing has been demonstrated in a number of viruses including SV5 (Thomas *et al.*, 1988), MV (Cattaneo *et al.*, 1989), SEN (Vidal *et al.*, 1990a), MuV (Paterson and Lamb, 1990, Takeuchi *et al.*, 1990), HPIV-2 (Ohigimoto *et al.*, 1990, Southern *et al.*, 1990) HPIV-4A and 4B (Kondo *et al.*, 1990), BPIV-3 (Pelet *et al.*, 1991), HPIV-3 (Galinski *et al.*, 1992), phocid distemper virus (Blixenkrone-Moller *et al.*, 1992) and porcine LPMV (Berg *et al.*, 1992). RNA editing activity results in the insertion of non-templated G nucleotides at specific sites during transcription of the P gene which causes a frame-shift, thus allowing access to alternative reading frames. These above viruses fall into two groups based on the way in which they express their P proteins. For MV, SEN, BPIV-3, HPIV-3 and phocid distemper virus, the unedited mRNA is translated to give the P protein whereas the edited mRNA accesses the V ORF. However, with MuV, SV5, HPIV-2 and 4, and LPMV, the P protein is expressed from an edited mRNA and the V protein is expressed from the unedited message. Interestingly the V protein is the most highly conserved of all the paramyxovirus and morbillivirus P gene products, particularly in a 55 amino acid domain which contains six conserved cysteine residues (Thomas *et al.*, 1988). The similarity of these V proteins to transcriptional cofactors such as zinc-finger proteins (Klug and Rhodes, 1987) and the HIV tat protein (Frankel *et al.*, 1988) has led to speculation that the V protein may be involved in RNA synthesis and available evidence supports this (Curran *et al.*, 1991a, 1991b).

In addition to the mechanism of RNA editing, many of these viruses contain alternative cistrons which are accessed by independent ribosomal initiation. The P

gene of SEN has been particularly well studied and has been found to give rise to 8 polypeptides by using both expression strategies (Curran *et al.*, 1991a, b). The SEN P gene has three ORFs, the P ORF, the C ORF and the V ORF (figure 3). The C ORF is expressed by independent ribosomal initiation at four separate initiation codons, one of which is a non-AUG codon (ACG), to give rise to the C, C', Y1 and Y2 polypeptides. The V ORF is expressed from an mRNA which contains a single G nucleotide insertion at a specific site and causes a switch from the P ORF to the V ORF at this position. The insertion of G residues at the insertion site is not always precise and the addition of two G residues rather than one leads to the transcription of an mRNA capable of expressing the W protein which represents a truncated form of P with two codons of the C ORF at its carboxyl-terminal end. The final polypeptide expressed from the SEN P gene is the X protein which is thought to be expressed from the P protein ORF by ribosomal initiation on either of two AUG codons located at least 1500 nucleotides from the 3' end of the P gene coding region. Which of the P, C, V or W mRNAs are responsible for expression of the X polypeptide and also the way in which scanning ribosomes are able to initiate at this codon when it is positioned so far from the 5' mRNA cap structure is unclear (Curran *et al.*, 1988b). The insertion of non-templated G residues is believed to occur by controlled polymerase stuttering in which the polymerase pauses on the P gene and the 3' end of the newly formed mRNA slips upstream of the template re-exposing an already copied region of the template which can then be re-transcribed into the nascent mRNA (Vidal *et al.*, 1990b). For each group of viruses, the precise number of Gs which correctly change the P and V ORFs are inserted with high frequency and the slippage model is able to explain this feature due to the stability of the miss-aligned intermediate which in the case of SEN occurs after a one base slippage and in the case of BPIV-3, occurs after a two base slippage. The expression of the P gene is further complicated in HPIV-3 and BPIV-3 in that a

(a) P/C mRNA



(b) V/C mRNA

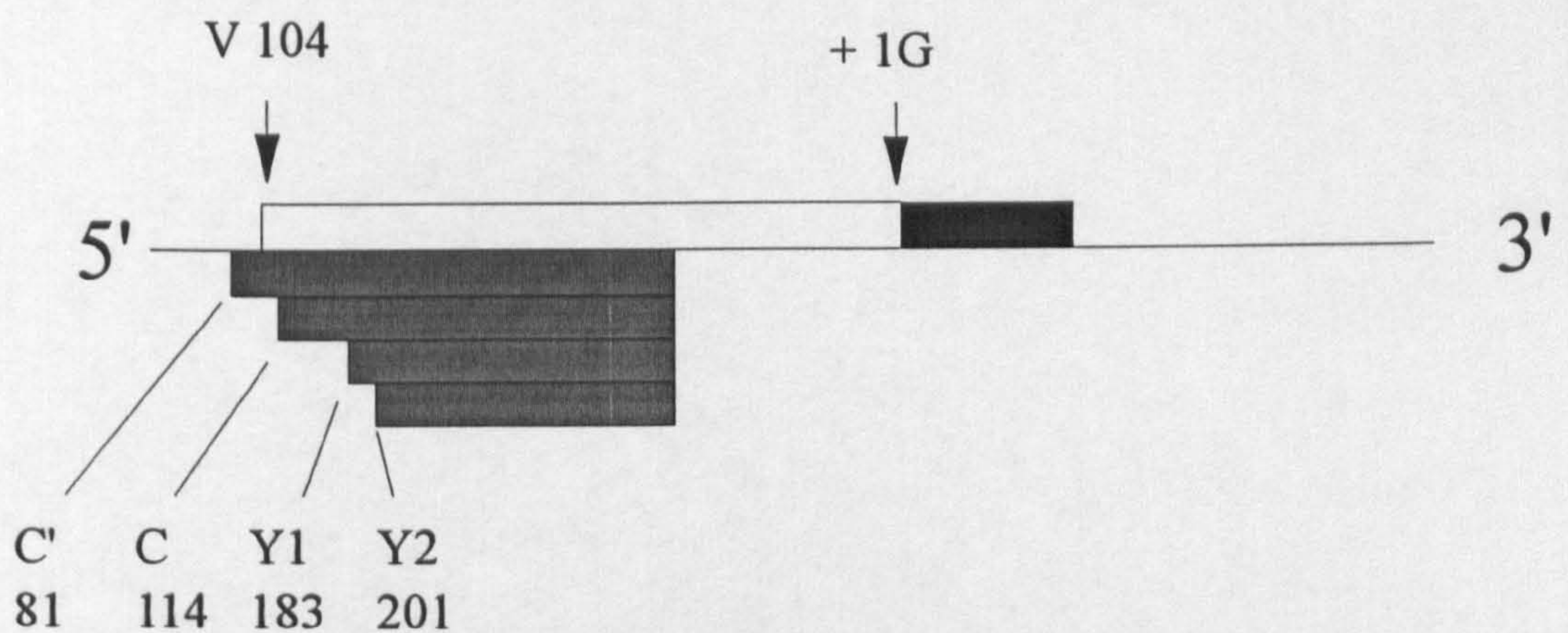


Figure 3. Protein coding regions of the SEN P/C and SEN V/C mRNAs (adapted from Kolakofsky *et al.*, 1991). The P, C, and V ORFs are shown as clear, shaded and black rectangles respectively. The nucleotide positions of several start codons and also the polypeptides they encode are indicated. The insertion site at which a single nucleotide is added in the synthesis of the V/C mRNA and which leads to a fusion of the P and V ORFs is also indicated. The X protein is thought to be expressed by initiation on AUG 1523. Addition of two G residues at the insertion site gives rise to an mRNA that encodes the W protein which represents a fusion of the P and the C ORFs.

fourth ORF, called the D ORF, exists which is in the same phase as the C ORF and which is accessed by insertion of 2 G residues at the insertion site (Pelet *et al.*, 1991). Although two slippage sites have been identified on the BPIV-3 P gene only the first one appears to be functional and access to all the P ORFs is achieved by removing the constraints which control the number of slippage rounds that can take place at this position. The C proteins expressed from the P gene of SEN and several of these other viruses seem to be present at very low abundances as structural components of purified virions (Yamada *et al.*, 1990), and since they can clearly associate with nucleocapsid components, much attention has been focused on assessing whether these polypeptides play a role in the regulation of the function of the viral polymerase.

1.4.4: The Large protein.

A cDNA clone representing the PVM L protein gene has not yet been obtained and this reflects the scarcity of L gene mRNA due to its location on the genome furthest from the postulated 3' polymerase binding site (figure 1). The PVM L protein migrates in SDS-PAGE with an observed molecular weight of 191 kDa (Cash *et al.*, 1979; Ling and Pringle, 1989a) which is similar to the observed molecular weight of the HRSV L protein which Cash *et al.*, (1977) determined to be approximately 200 kDa and Ling (1989) determined to be 190 kDa.

The nucleotide sequence of the HRSV L protein was recently determined (Stec *et al.*, 1991). The deduced amino acid sequence of the HRSV L protein is 2165 amino acids in length and has a calculated molecular weight of 250,226 Da which is in close agreement with the observed molecular weight values given above, given the inherent inaccuracies of the gel system used. The 22K gene of HRSV contains the gene start signal for the L gene (Collins *et al.*, 1987). The two genes overlap by 68 nucleotides and whilst the L gene overlap contains the beginning of the L ORF, the sequence of the 22K gene in the overlap region is non-coding. A consequence of this overlap is that L

gene start is followed 68 nucleotides later by the 22K gene stop signal, and thus in order for the L gene to be expressed, this stop sequence must be ignored. The presence of the gene end sequence within the L gene leads to drastic attenuation of L gene transcription. The PVM 22K and L genes do not appear to overlap in this way (P. Chambers, unpublished data).

By analogy with other similar viruses, the pneumovirus L protein is believed to represent the major component of the viral RNA dependent RNA polymerase which is responsible for the synthesis of viral mRNAs, full-length replicative positive sense genomes and progeny vRNA genomes. Also by analogy, the pneumovirus L protein would be expected to possess enzymatic activities of mRNA capping, methylation and polyadenylation and the ability to act as a protein kinase (Hamaguchi *et al.*, 1983; Hunt *et al.*, 1984; Sanchez *et al.*, 1985). Since the HRSV L protein sequence was reported, several observations have been made which support the validity of these functional assignments for the pneumovirus L protein. The HRSV L protein is very similar in length to the L proteins of other non-segmented and segmented negative stranded viruses, which are remarkably constant in length being between 2109 and 2238 amino acids long (Stec *et al.*, 1991). Poch *et al.*, (1989) identified 4 regions within a variety of RNA dependent RNA and DNA polymerases that were proposed to be involved in polymerase function. These 4 domains are present in the HRSV L protein which suggests distant relatedness and common function. Statistically important sequence similarity was detected between the HRSV L protein and several other non-segmented negative stranded RNA viruses. In descending order, the most similar were the L proteins from PIV-3, MV, SEN, NDV and rabies virus. The L protein is only the third pneumovirus protein, after the HRSV F protein (Spriggs and Collins, 1986) and the PVM N protein (Barr *et al.*, 1991) to share some degree of amino acid relatedness with other members of the family *Paramyxoviridae*, although the homology demonstrated by the L protein is by far the most significant.

1.4.5: The Matrix protein.

The Matrix protein gene of PVM is 927 nucleotides in length and has the capacity to encode a polypeptide 257 amino acids in length (P. Chambers, pers. comm.). The predicted molecular weight of this polypeptide is 28,312 Da, which closely agrees with the observed molecular weight of the putative PVM M protein in SDS-PAGE (27.3 kDa; Ling and Pringle, 1989a). The PVM M protein shows a good degree of amino acid identity with the M protein of HRSV (41.3%; Satake and Venkatesan, 1984), however no identity has been detected between the PVM M protein and the M proteins of any other *Paramyxoviridae* family members. The PVM M protein is both extremely basic and extremely hydrophobic. It has been suggested that a hydrophobic domain located in the carboxyl-terminal region of the HRSV M protein, which is also present in the PVM M protein, may be responsible for interaction with the cell membrane (Collins, 1991). Reports concerning pneumovirus M protein phosphorylation are conflicting; Gruber and Levine (1985) believed that the HRSV M protein was not phosphorylated whereas studies by Lambert (1988) suggested that it was. The M protein expressed in PVM-infected BSC1 cells is not thought to be phosphorylated (Ling and Pringle, 1989a).

The pneumovirus M protein has not been functionally characterised, and so functions are inferred from the more extensively characterised counterparts in the *Paramyxoviridae* and *Rhabdoviridae* families. The M proteins of many *Paramyxoviridae* family members are believed to be involved in functions relating to maturation and assembly of the virion, and it has also been suggested that the M protein may be able to regulate RNA synthesis by interacting with active nucleocapsid structures. The M protein is considered to form an essential bridge between the assembled nucleocapsid and the cytoplasmic tail of the viral glycoprotein, and this supposition is strengthened by studies that have indicated that the M protein of *Paramyxoviridae* family members is able to interact with the nucleocapsid structure,

the viral glycoproteins and also itself (Peeples, 1991). In addition, the M protein is able to interact with a variety of cellular components such as membrane lipids and cytoskeletal elements which may be involved in the assembly of virus particles, and also in virus budding.

The association of the NDV M protein with plasma membranes was initially revealed by cell-fractionation experiments (Schwalbe and Hightower, 1982). Faaberg and Peeples (1988) demonstrated that the ability of NDV M protein molecules to associate with lipid membrane structures in the form of liposomes was not governed by forces of an electrostatic nature, but that the interaction probably arose through the presence of a hydrophobic domain on the M protein. Charge independent association of M protein and lipid membranes was previously shown to occur with the M protein of influenza virus (Bucher *et al.*, 1980). In contrast to these results was the finding that the interaction between the VSV M protein and lipid membranes was dependent to some degree on electrostatic attraction since M protein only binds negatively charged liposomes (Zakowski *et al.*, 1981). The nature of the interaction between VSV M protein and the cell membrane still remains unclear. The VSV M protein is believed to associate with the membrane of Hela cells in the absence of all other virus encoded components and during the interaction with the membrane, part of the M protein is thought to be extended into the membrane interior (Chong *et al.*, 1993). Experiments by Li *et al.*, (1993) using a baculovirus expression system, have indicated that the VSV M protein has the intrinsic property of promoting invagination of a circumscribed region of the host cell cytoplasmic membrane, leading to the liberation of liposome structures. This function was found to be solely directed by the M protein, a finding contradicted by Pattnaik and Wertz (1991) using a vaccinia virus T7 expression system, who did not observe budding when M protein was expressed alone. The ability of VSV M protein to independently bind to membranes differs from the situation in SEN where the M protein was found to be unable to associate with membranes when expressed on its own (Sanderson *et al.*, 1993).

There is much evidence to suggest that the M protein is able to associate with viral glycoproteins. In VSV, the M protein has been shown to be physically attached to G protein in virions through chemical cross-linking experiments (Dubovi and Wagner, 1977). The VSV M protein is believed to associate with G protein in a specific and reversible manner which is a necessary requirement to account for the proposed role of the M protein in virus assembly (Lyles *et al.*, 1992). These workers also revealed that the binding of M protein to G was enhanced when the M protein was supplied in the form of N-M complexes, supporting the assembly model in which the M protein brings the nucleocapsid and glycoprotein laden membrane together in the process that leads to budding of the virus particle. In either paramyxovirus or morbillivirus systems, demonstration that a definitive physical relationship exists between M and G proteins has not yet been reported, although the observation that cellular location of membrane bound M protein appears to be determined by the location of the SEN glycoproteins suggests that these protein species do associate (Sanderson *et al.*, 1993).

Whereas the ability of the paramyxovirus M protein to bind to glycoproteins is somewhat unclear at present, chemical cross-linking experiments have unequivocally shown that it is capable of interacting with itself (Nagai *et al.*, 1978). In addition, electron micrographs of preparations of purified M protein have revealed the ability of M protein to associate and form tube structures composed of a two-dimensional crystalline array of parallel strands wound into a helix, reminiscent of structures found on the cytoplasmic surface of infected cell plasma membranes. Interaction of the paramyxovirus M protein with the N protein is also well established. Homann *et al.*, (1991) demonstrated specific binding between the SEN M and N proteins using a protein blotting technique. These workers also demonstrated that the M protein was capable of association with an unidentified cellular component, a function also suggested by Sanderson *et al.*, (1993).

The SEN M protein has been shown to inhibit, by 90%, the ability of purified nucleocapsids to transcribe RNA (Marx *et al.*, 1974) and the inhibitory effect of M protein on transcription of VSV genomes is well documented (De *et al.*, 1982; Black *et al.*, 1992; Black *et al.*, 1993; Li *et al.*, 1993). The M protein is able to condense the nucleocapsid into transcriptionally inactive forms, the inactivity presumably being due to structural hindrance of the polymerase as it attempts to pass along the template (De *et al.*, 1982). It has been suggested that the VSV M protein associates with the nucleocapsid by its amino-terminal end (Shipley *et al.*, 1988).

It has been suggested that defects in the M protein may be responsible for establishing MV persistence. M proteins isolated from a patient with SSPE were found to be unable of associating with nucleocapsids (Hirano *et al.*, 1992) and the M gene of a MV isolate taken from a patient afflicted with SSPE was found to contain an amber mutation which prevented expression of functional M protein (Cattaneo *et al.*, 1986).

1.4.6: The 22K (M2) protein.

The 22K gene of PVM is 868 nucleotides long and has the potential to encode a polypeptide 176 amino acids in length. Using two dimensional SDS-PAGE, Ling and Pringle (1989a) identified a basic polypeptide having a molecular weight of between 23 kDa and 24 kDa which they suggested was the PVM 22K species. The PVM 22K protein shows considerable amino acid identity to the 22K proteins of HRSV, BRSV and TRTV, and these three polypeptides have similar hydropathy plots which indicates a degree of structural conservation over regions where amino acid identity is not strong (Yu *et al.*, 1992b; Collins and Wertz ., 1985; Zamora *et al.*, 1992). The pneumovirus 22K proteins have not been functionally characterised, and since the 22K proteins are unique to the pneumoviruses, no possible functions can be inferred from other related viruses. Fluorescent antibody studies have suggested that, in HRSV-infected cells, the HRSV 22K protein associates with the N and P proteins (Routledge

et al., 1987). In addition, the N and 22K proteins of BRSV have been shown to associate *in vitro* using a protein blotting technique (Samal *et al.*, 1993). The interaction of 22K and N proteins both in HRSV-infected cells and also in cells transfected with vaccinia vectors was demonstrated by Garcia *et al.*, (1993) using immunofluorescence and immunoprecipitation studies. In addition to the identification of internally located 22K protein, the 22K species has been shown to be present within the cell membrane of HRSV-infected cells by salt and detergent dissociation studies (Huang *et al.*, 1985). Thus, like the M protein, the 22K protein seems to be found in association with both nucleocapsid structures and also the cell membrane, suggesting that, like M protein, it may be involved in virion assembly. It has been suggested that the functions fulfilled by the M protein of other non-segmented negative sense RNA viruses has been split between the pneumovirus M and 22K proteins (Collins, 1991).

1.4.7: The 1B and 1C proteins.

The genes which encode these two proteins are located closest to the 3' end of the PVM genome. Given a single entry site for transcription by the pneumovirus RNA dependant RNA polymerase at the 3' end of the genome, and attenuation of transcription at successive intergenic regions, this location would suggest high levels of transcription of mRNA species corresponding to these genes (figure 1 and section 1.6.4). This suggestion is supported by the high proportion of 1B and 1C mRNA species cloned as cDNA, reported during the molecular cloning of PVM (Chambers *et al.*, 1990a), and also the high levels of the 1B and 1C proteins in PVM-infected BSC1 cells (Cash *et al.*, 1979; Ling and Pringle, 1989a).

The 1B and 1C genes are 410 and 571 nucleotides in length respectively. The 1C gene polypeptide product is predicted to be 113 amino acids long and have a molecular weight of 12,906 Da. This is in close agreement with the observed molecular weight for this species in PVM-infected BSC1 cells which is between 14 and 16 kDa,

depending on the SDS-PAGE system used (Cash *et al.*, 1979; Chambers *et al.*, 1990a; Ling and Pringle, 1989a). The 1B gene polypeptide product is 156 amino acids in length and is predicted to have a molecular weight of 18,234 Da which is close to the molecular weight of this species determined by SDS-PAGE, which varies between 17 kDa and 20 kDa depending on the conditions of electrophoresis (Cash *et al.*, 1979; Ling and Pringle, 1989a; Chambers *et al.*, 1990a). Interestingly the 1C gene has an extensive second open reading frame which is accessed by the third AUG codon from the 5' end of the mRNA and is capable of encoding a polypeptide 69 amino acids in length. In addition, the first AUG codon of the 1C gene is followed by a second AUG codon only nine nucleotides downstream. Neither of these AUGs are in good context for initiation (Kozak, 1986a, 1986b) but it is possible that they are both used since two closely migrating forms of the 1C polypeptide are seen in PVM-infected BSC1 cells (Ling and Pringle, 1989a). Perhaps the poor initiation context for these two in-frame AUG codons allows scanning ribosomes to access the second ORF.

The 1B and 1C genes of PVM are not detectably homologous to the 1B and 1C genes of HRSV, or any other HRSV gene. A comparison of amino acid homologies between the 1B and 1C gene products and their HRSV counterparts was more successful in that a moderate level of homology (29% identity) was identified over a 35 amino acid carboxyl-terminal segment of the PVM and HRSV 1B polypeptides (Chambers *et al.*, 1991).

The 1B and 1C polypeptides of PVM or HRSV have not been functionally characterised although due to their absence in the virion they appear not to be structural, and this observation invites the suggestion that these polypeptides may be involved in replication and transcription of the genome.

1.4.8: The fusion protein.

The F gene of PVM is 1657 nucleotides in length and encodes a polypeptide 537 amino acids in length (Chambers *et al.*, 1992). From the deduced amino acid sequence, the predicted molecular weight of the F protein is 59,374 Da. Hybrid arrested translation experiments identified the F gene protein product to have a molecular weight of only 49,000 Da (Chambers *et al.*, 1990a). The PVM F protein shows approximately 40% overall amino acid identity with the HRSV and TRTV F proteins (Chambers *et al.*, 1992). The amino acid sequence identity between the BRSV and HRSV F proteins is 80% and 81%, depending on HRSV subgroup, and this figure is slightly lower than for the identity between the F proteins of the two HRSV subgroups which is 90% (Walravens *et al.*, 1990). In common with other PVM proteins, comparisons between the PVM F protein and the F proteins of other paramyxovirus and morbillivirus members revealed only a low (10%) overall amino acid identity (Chambers *et al.*, 1992). A slightly higher degree of amino acid identity exists between the F proteins of HRSV and several other paramyxoviruses (Spriggs and Collins, 1986). Despite the lack of amino acid identity between the F proteins of viruses belonging to these groups, structural analyses has revealed the presence of various secondary structural elements that were well conserved between the F proteins of all morbillivirus and paramyxovirus members (Chambers *et al.*, 1992). This observation is pertinent as all the *Paramyxoviridae* family members studied to date possess F proteins that share characteristics presumably relating to a common function. These common characteristics as well as the various structural motifs with which these characteristics are associated are described below.

The F proteins of *Paramyxoviridae* family members are synthesised as a full-length inactive precursor (F_0) which is then cleaved by cellular proteases to yield a functional moiety composed of F_1 and F_2 subunits which are linked by a disulphide bond. The cleavage site responsible for generating the two PVM F protein subunits has been suggested to lie between amino acids 101 and 102 by analogy with the known

cleavage site in the F protein of HRSV (Chambers *et al.*, 1992; Elango *et al.*, 1985). The cleavage releases a 101 amino acid long F₂ subunit and a 436 amino acid long F₁ subunit. The F₁ subunit of PVM contains two potential acceptor sites for N-linked glycosylation whereas the F₂ subunit contains none. The distribution of potential glycosylation sites in HRSV is somewhat different with all but one of the sites (a total of 5 or 6, depending on the strain) being located in the F₂ subunit.

In common with F proteins of other *Paramyxoviridae* family members, the PVM F protein has three highly hydrophobic regions which have been assigned separate functions. Firstly, located at the extreme amino-terminal of the F₂ subunit is a hydrophobic signal peptide that is responsible for the cotranslational translocation of the F protein through the rough endoplasmic reticulum and which is cleaved off during or after translocation. The second region is the fusion related domain which is located at the amino-terminus of the F₁ subunit generated following cleavage. This region is thought to mediate fusion between both virus particle and cell membrane and also between adjacent infected and uninfected cells by localised disruption of the lipid membrane. The third region which is the most hydrophobic of all is the putative transmembrane region which is located at the carboxyl-terminus of F₁. These three F protein regions are diagrammatically represented in figure 4.

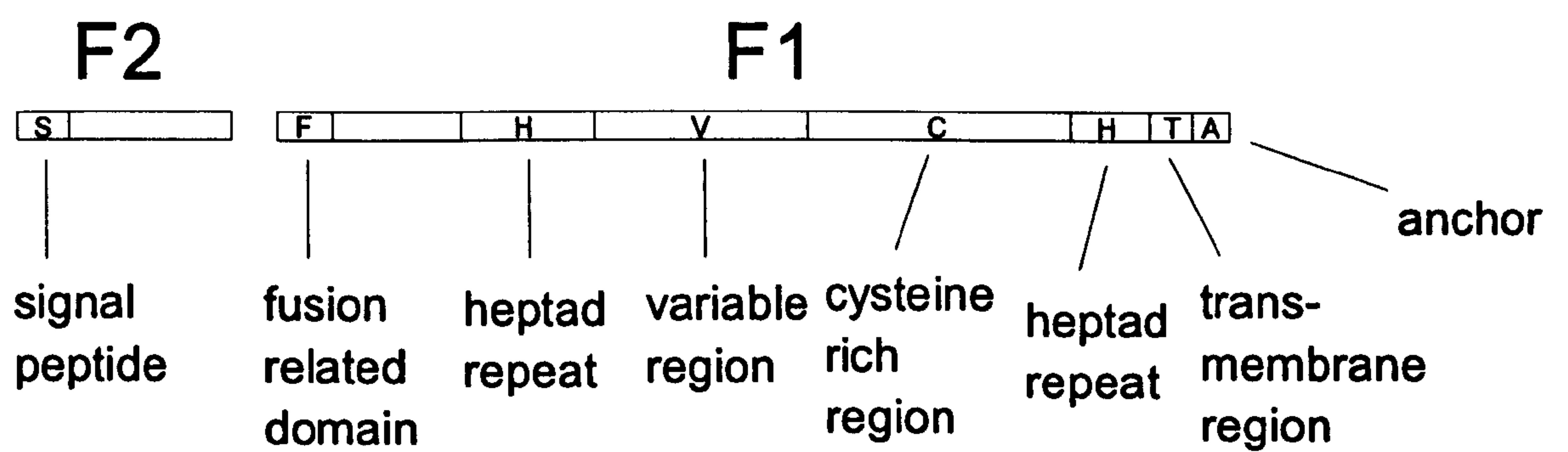


Figure 4. Schematic representation displaying structural and functional features that are common to the F1 and F2 subunits of the F proteins of *Paramyxoviridae* family members.

In addition to these conserved regions of hydrophobicity, several other secondary structural features shared by the F proteins of *Paramyxoviridae* family members have been identified (Chambers *et al.*, 1992). The amino-terminal region of F₂ is thought to be composed of β -sheet and turn or coil structures. The carboxyl-terminus of F₂ is thought to form an amphipathic α -helical structure. These two regions of F₂ are separated by a region that contains the only cysteine of F₂, the presence of which is essential in order for cross-linking to F₁. Not surprisingly, this cysteine is conserved in all paramyxovirus F proteins. The amino-terminus of F₁ contains the previously mentioned hydrophobic fusion domain. Located on the carboxyl-terminal side of this is a region of secondary structure known as a heptad repeat. The heptad repeats are sequences of amino acids arranged with a seven residue periodicity in which positions 1 and 4 are predominantly occupied by small hydrophobic or neutral residues. These heptad repeat regions suggest the presence of extended α -helical conformations. On the carboxyl-terminal side of the heptad repeat is one region which may form helical structures, and another which displays relative variability between the F proteins of the *Paramyxoviridae* family members. On the carboxyl-terminal side of the variable region is a cysteine rich domain where the location of seven or eight cysteines are conserved between all paramyxovirus and pneumovirus genus members (Shioda *et al.*, 1986). Between this region and the aforementioned hydrophobic transmembrane region, located at the carboxyl-terminal end of the F₁ subunit, is another region of heptad repeat which has been likened to the leucine zipper motif described by Buckland and Wild (1989). By specific alteration of residues in the hydrophobic amino-terminal domain of paramyxovirus F proteins it has been shown that amino acid changes that increase the helical content of the transmembrane region increase the cell-fusion activity of mutant F proteins (Horvath and Lamb, 1992).

The SEN F protein has been isolated as a tetramer and due to the known similarities that exist between the paramyxovirus and pneumovirus F proteins it is possible that the PVM or HRSV F proteins may exist in a similar form. The PVM F protein may form multi-subunit assemblies constructed around the region where the heptad repeat and α -helical coils are located, and additionally, the highly hydrophobic nature of the F protein may aid hydrophobic contact between individual subunits (Chambers *et al.*, 1990c).

Investigations into the mechanism responsible for cleavage of the F protein into the two subunits determined that amino acid changes in the F₁ subunit were responsible for the loss of F₀ cleavage, thus alteration of one region of the F protein can influence changes in another that is spatially separate (Anderson *et al.*, 1992).

1.4.9: The attachment protein

The PVM G gene is 1330 nucleotides long and encodes a polypeptide 363 amino acids in length. The 3' end of the PVM strain 15 G gene contains a small ORF beginning at an AUG at nucleotide position 83. This ORF overlaps with the large second ORF which is initiated from an AUG codon at nucleotide position 183 and is believed to express the G protein. In PVM strain 3666, the first and second ORFs have been merged and expression of the G protein is believed to be initiated from the AUG codon at nucleotide position 83 (J. Rhandawa, pers. comm). The predicted molecular weight of the G protein polypeptide moiety is 39,836 Da. In PVM-infected BSC1 cells, two forms of the PVM G protein are present having observed molecular weights of 76.4 and 62 kDa, and these are designated G₁ and G₂ (Ling and Pringle, 1989b). These polypeptides were both shown to contain common protein sequence by peptide mapping. In pulse-chase labelling experiments, a temporal relationship was also demonstrated between G₁ and G₂, with G₂ apparently being derived from G₁. A G protein related polypeptide species which is analogous to the PVM G₂ protein is also

seen in HRSV-infected cells. This form of G, which accounts for about one-sixth of all G protein forms is soluble and is secreted from infected cells (Hendricks *et al.*, 1987, 1988). This G protein species is cleaved co or post-translationally at either amino acid position 65 or 74, thus separating the hydrophobic amino-terminus from the rest of the protein (Hendricks *et al.*, 1988). However, it has recently been reported that the soluble form of G may be expressed by internal initiation from the AUG codon corresponding to methionine residue 48 (Roberts *et al.*, 1993). This amino-terminally truncated form of G is believed to be further amino-terminally truncated by proteolytic cleavage at the positions previously identified by Hendricks *et al.*, (1988).

The large discrepancy between the predicted and observed molecular weights of the G protein is due to extensive glycosylation. Treatment of the 76.4 kDa G₁ polypeptide with glycopeptidase F, an agent known to specifically remove N-linked carbohydrates, gave rise to a polypeptide with a molecular weight of 58.4 kDa. Further treatment of the 58.4 kDa species with a range of agents capable of removing O-linked carbohydrates gave rise to a polypeptide with an observed molecular weight of 39.6 kDa, and this species was believed to represent the polypeptide moiety of the PVM G protein (Ling and Pringle, 1989b). Similar reductions in observed molecular weight were obtained for the G₂ polypeptide, and the purportedly unglycosylated G₂ species was calculated to have a molecular weight of between 37.2 kDa and 44.8 kDa. The haemagglutination activity of PVM is associated with the G protein since anti-G protein monoclonal antibody strongly inhibits the activity (Ling and Pringle, 1989b).

At 298 or 292 amino acids, depending on subgroup, the HRSV G protein is considerably shorter than the PVM G protein and it has been suggested that the additional sequences of the PVM G protein may be responsible for the haemagglutination activity which the HRSV G protein lacks. The overall amino acid homology that exists between the PVM and HRSV G proteins is poor (J.Rhandawa, personal communication). The HRSV G protein is also heavily glycosylated, with an

observed molecular weight change of between 8 and 12 kDa being due to N-linked sugars and over 40 kDa being due to O-linked sugars, the precise degree of glycosylation varying from strain to strain due to a variation in the precise number of N and O-linked glycosylation acceptor sites. The HRSV G protein is glycosylated to a greater degree than the PVM G protein, with carbohydrate increasing the G protein molecular weight by approximately 50 kDa compared with approximately 20 kDa for the PVM G protein. The large molecular weight difference attributed to O-linked sugars may however be an artefact due to reduced SDS binding on account of the O-linked sugars (Collins and Mottet, 1992). Investigations by several groups have determined that the N-linked sugars are added co-translationally whereas the O-linked sugars are added post-translationally (Ferne *et al.*, 1985; Wertz *et al.*, 1989). There is no sequence homology between the pneumovirus G proteins and the G proteins of other *Paramyxoviridae* family members, and the only structural feature that these proteins share is that they expose their amino-terminal ends on the cytoplasmic side of the membrane into which they are inserted (type II membrane orientation; Haepfle *et al.*, 1989).

As for PVM, the first initiation codon of the HRSV G ORF is followed soon after by a termination codon, and so initiation on the second in-frame AUG codon is believed to give rise to the HRSV G protein. In both viruses, this arrangement of initiation codons is likely to down-regulate expression of the G protein as scanning ribosomal 40S subunits will probably initiate at the 5' proximal AUG. Interestingly, the first AUG of the TRTV G gene is not followed soon after by a termination codon, and thus expression of TRTV G is not down-regulated. However, similar down-regulation of TRTV G protein expression relative to other TRTV proteins may occur as a consequence of the position of the G gene on the TRTV genome, since in TRTV the G gene is further from the 3' proximal promoter region (Yu *et al.*, 1992c) and therefore is less frequently transcribed.

Attachment of HRSV to cells was shown to be inhibited by mono-specific anti-G protein antisera thus implicating G as the attachment protein (Levine *et al.*, 1987). The receptor responsible for G protein attachment has still not been identified although it appears to be a ubiquitous cell membrane component since Walsh *et al.*, (1984) were unable to saturate the receptor using purified G protein. The similarities that exist between G proteins of HRSV and PVM suggest that in PVM, G is also the attachment protein. The G proteins of PVM and of strains from both HRSV subtypes share certain structural elements to which functions have been defined. The G protein amino-terminus contains a single major hydrophobic domain which for several reasons is believed to specify the functions of membrane insertion and anchorage. Firstly, a peptide expressed from a recombinant vaccinia vector representing the amino-terminal 71 amino acids of the HRSV G protein was transported to, and inserted and anchored into, the cell membrane (Olmsted *et al.*, 1989). Secondly, expression of HRSV and PIV-3 attachment proteins in which the amino-terminal hydrophobic regions had been interchanged were processed in an authentic manner (Collins, 1991). Finally, peptide sequence analysis has revealed that the soluble form of the HRSV G protein which is not inserted into the membrane (analogous to the PVM G₂ protein) is lacking the amino-terminal hydrophobic element of the full-length G protein (Hendricks *et al.*, 1988). It has been suggested that the amino-terminal hydrophobic element of the G protein can be divided into two domains, one of which is dispensable for membrane insertion (Lichenstein *et al.*, 1991). The region of the G protein on the carboxyl-terminal side of the membrane anchor is believed to represent the ectodomain. This region shows lower amino acid conservation between both inter-subgroup and intra-subgroup HRSV isolates than is exhibited between the relatively conserved transmembrane and cytoplasmic regions. The region of the G protein bounded by amino acids 101-133 which lies immediately carboxyl-terminal to the transmembrane region shows greatest intra-subgroup variation and for this reason is known as the hypervariable region. This portion of the G protein is separated from another variable

region located at the carboxyl-terminal third of the G protein by a small conserved region between amino acids 147 and 207 (Cane *et al.*, 1991). The observation that 77 of the 91 potential O-linked glycosylation sites and all of the N-linked sugar attachments sites are located on the carboxyl-terminal side of the membrane anchor supports the suggestion that this is the proposed ectodomain of the G protein (Sullender and Wertz, 1991).

The presence of O-linked glycosylations on the G proteins is a feature that within the *Paramyxoviridae* family of viruses, is unique to the pneumoviruses. The ectodomains of the G proteins of both HRSV subgroups contain over 90 potential acceptor sites for O-linked glycosylation clustered predominantly into two domains, one at the carboxyl-terminus and one adjacent to the membrane anchor, however the exact location of these sites is not well conserved (Olmsted *et al.*, 1989). The absence of either the N-linked or the O-linked glycosylations reduces the level to which the HRSV G protein is expressed on the cell surface to about 50% that of fully glycosylated G protein (Wertz *et al.*, 1989). Completely unglycosylated G protein is also expressed on the cell membrane but this time at a level only 10% that of mature G. By using three inhibitors specific for different steps of exocytosis, Collins and Mottet (1992) determined that initial addition of O-linked sugars begins in the trans-Golgi compartment with final maturation occurring in another compartment of unknown location. The HRSV G protein is believed to associate to form homo-oligomers (Lambert, 1988) and this association seems to occur when the G protein is devoid of N-linked sugars. Removal of N and O-linked oligosaccharides from the G protein drastically reduces the potential infectivity of virus particles which suggests that the oligosaccharide component of the G protein exerts considerable influence on the attachment process (Lambert, 1988). The presence of O-linked sugars is not thought to affect the folding and oligomerisation of the G protein, however, the presence of O-linked sugars on the G protein may be important to confer stability of an already formed oligomeric structure (Collins and Mottet, 1992).

The extensive glycosylation to which the G protein is subjected may serve to alter its antigenicity, and thus evade the host immune response. Out of a panel of 18 monoclonal antibodies raised against the HRSV G protein, only five recognised unglycosylated G protein whilst the majority recognised G protein glycosylated only with O-linked sugars (Palomo *et al.*, 1991). Of these 18 antibodies only those that recognised unglycosylated G protein were able to neutralise HRSV infectivity to any degree (Garcia-Bareno *et al.*, 1989). G protein monoclonal antibodies are characteristic in their poor neutralising ability and this may reflect the heterogeneity of the G protein population, such that a large non-neutralisable fraction of virus is often present even from a population of virus that is genetically homogeneous. Characterisation of neutralisation resistant mutants of HRSV has highlighted the heterogeneity of the G protein population (Garcia-Barreno *et al.*, 1990, 1992; Rueda *et al.*, 1991). Escape mutants which were resistant to anti-G protein monoclonal antibodies arose by way of a mechanism which generated either point mutations, premature stop codons or frameshift mutations in the G gene. Many of the mutant G proteins were caused by addition or removal of adenosine residues in A runs of the corresponding G gene probably by polymerase slippage during genome replication. The mechanism is probably similar to that which allows access to alternative reading frames in the P genes of many *Paramyxoviridae* family members. Interestingly, the majority of mutations in the G proteins of these escape mutants affected the carboxyl-terminal region of the protein and this indicated that the carboxyl-terminal of G can accommodate sequence changes and still remain functional. Naturally occurring isolates of HRSV also show considerable variation in the sequence of G proteins (Cane *et al.*, 1991; Sullender *et al.*, 1991).

1.4.10: The SH (1A) protein.

The SH protein gene of PVM is 390 nucleotides long and can encode a polypeptide 92 amino acids in length with a predicted molecular weight of 9,582 Da. The nucleotide sequence of the SH protein gene revealed the presence of four N-linked glycosylation sites, with two located at the amino-terminal end and two located at the carboxyl-terminal end of the protein. The predicted amino acid sequence of the TRTV SH protein has revealed the presence of a potential N-linked glycosylation site and this site is thought to be used (Ling *et al.*, 1992). The SH polypeptide has not yet been unequivocally identified in PVM-infected cell material although Ling and Pringle (1989b) identified a 12 kDa glycosylated polypeptide that could have represented this species. The SH protein of HRSV has been shown to be synthesised in four different monomeric forms with molecular weights of 4,500 Da, 7,500 Da, 13,000-15,000 Da and 21,000-30,000 Da (Olmsted and Collins, 1989). The 7,500 Da species appears to represent the full-length unglycosylated form of SH. The 4,500 Da form of SH is also unglycosylated and is believed to arise through translational initiation from an AUG codon located 22 codons downstream of the AUG codon responsible for the full-length SH species. The 13,000-15,000 Da species represent N glycosylated forms of SH, and a proportion of these are further modified by the addition of polylactosaminoglycan to the N-linked carbohydrate to form the 21,000-30,000 Da species. All four of these forms are found associated with the cell membrane of HRSV-infected cells (Olmsted and Collins, 1989). In addition, the 7,500 and the 21,000-30,000 Da species were found in purified virions suggesting that these two forms of SH are structural (Collins, 1991). Amino acids 23 to 41 of the SH protein are responsible for membrane anchorage which suggests that the amino-terminal 22 amino acids are either membrane embedded or exposed to the cell interior (Collins and Mottet, 1993). Interestingly, comparison of the amino acid sequence of the SH proteins of HRSV strain A2 and 18537 isolates (from subgroups A and B respectively) showed a high degree of conservation in the membrane and cytoplasmic region of SH (88%), but more divergence in the carboxyl-terminal region (50%; Collins *et al.*,

1990). Chemical cross-linking experiments have suggested that the full-length unglycosylated SH protein is able to form homodimers, homotrimers, homotetramers and homopentamers (Collins and Mottet, 1993). An additional species with a molecular weight of 180 kDa was also identified and believed to represent an oligomer composed of four or five cross-linked pentameric subunits. The function of the pneumovirus SH protein is presently unknown.

1.5: PNEUMOVIRUS MOLECULAR BIOLOGY.

1.5.1: The pneumovirus genome.

The molecular cloning of PVM has been achieved and this has provided the means by which to rapidly further the understanding of this virus (Chambers *et al.*, 1990a). Nine genes have presently been identified on the PVM genome and their nucleotide sequences have been determined. Coding assignments have been made for these genes by both hybrid-arrested translation experiments, and also by comparison of the nucleotide sequence with HRSV genes for which assignments have already been made. The order in which the genes are arranged on the PVM genome was confirmed by determination of the nucleotide sequence of the intergenic regions in polycistronic cDNA clones. The gene order of PVM is identical to the gene order of HRSV. Like HRSV, it is believed that the PVM genome consists of a total of 10 genes, and the remaining PVM gene that has not been identified is believed to correspond to the large, or L, protein of HRSV which has been identified as a component of the viral RNA dependent RNA polymerase. The genome of HRSV has been shown to be of negative sense (Huang and Wertz, 1982) and by analogy, the genome of PVM is also believed to be of negative sense. The HRSV genome has been completely sequenced and has been determined to be 15,222 nucleotides in length (Collins, 1991). At the extreme 3' and 5' termini of the HRSV genome are the leader and trailer sequences respectively (Mink *et al.*, 1991) and regions which correspond to these are expected to be present at the 3' and 5' ends of the PVM genome. The leader and trailer regions are involved in transcription and replication of the viral genome and this is discussed in sections 1.5.2 and 1.6.4. The recent molecular cloning of TRTV has revealed that the order of genes on the TRTV genome is different to that of the PVM and HRSV genomes (Yu *et al.*, 1992c; Ling *et al.*, 1992). In TRTV, the SH and G genes have been moved downstream of the F and 22K genes, and this is shown diagrammatically in figure 1.

The different order of genes on the TRTV genome represents a fundamental divergence of this virus from the other pneumoviruses.

1.5.2: The leader and trailer regions.

By analogy with other *Paramyxoviridae* and *Rhabdoviridae* family members, the sequences at the 3' and 5' ends of the pneumovirus genome (the leader and trailer) are believed to harbour signals involved in pneumovirus RNA synthesis. The HRSV leader region is 44 nucleotides long and, as with other negative sense single stranded RNA viruses, is thought to encode at least two distinct nucleotide signals (Mink *et al.*, 1991). These are the regions responsible for binding the polymerase complex prior to RNA synthesis and a site responsible for initiating encapsidation of nascent, negative sense genome copies (vRNAs). In terms of nucleotide sequence, the HRSV leader region is distinct from the leader regions of other negative sense single stranded RNA viruses. The HRSV leader shares nucleotide identity between other members of the *Paramyxoviridae* family only at the first two nucleotide positions, and in only the first three or four nucleotide positions of the leader sequences belonging to rhabdoviruses, VSV and rabies virus, respectively (Mink *et al.*, 1991). At 155 nucleotides, the HRSV trailer region is much longer than any other equivalent sequence determined to date. The degree of nucleotide homology that exists between the HRSV trailer and the other trailer regions is similar to the level of homology that was detected between their respective leader regions. In common with other viruses from related groups, the HRSV leader and trailer regions show complementarity, and in the case of HRSV this extends to 21 of the first 26 nucleotides. The complementarity is essential for the RNA synthetic procedures since the 5' trailer of the (-) vRNA encodes the 3' leader of the positive sense genome copy (vcRNA) which is thought to be functionally analogous to the 3' leader of the vRNA. Many functional aspects of the leader and trailer regions

have been elucidated for VSV and to a lesser extent the paramyxoviruses SEN and NDV and these functions are discussed in sections 1.6.4 and 1.6.5.

1.5.3: Gene start and end sequences.

During HRSV transcription the polymerase initiates gene transcription at the gene start signal (Collins *et al.*, 1986). The HRSV gene end signals are believed to cause the viral polymerase to stall and, by reiterative copying of short poly(U) tracts, synthesise a poly(A) tail. It is believed that no other gene sequences other than those identified as gene start and gene end sequences are able to act as structural signals since these are the only regions that display conservation between HRSV subgroups (Collins, 1991 and references therein). The gene start and gene end sequences of nine PVM genes have been determined and these sequences, along with the overall start and end consensus sequences, are shown in table 2. The gene start and gene end sequences for HRSV are also shown. Although the HRSV and PVM consensus sequences are not identical, they are clearly related, and for this reason the PVM gene start and end sequences are expected to be functionally analogous to the HRSV sequences. The PVM gene start and stop sequences are fairly well conserved between genes, and this contrasts to HRSV in which the gene starts are well conserved but the gene ends show considerable variation. Interestingly, many of the PVM genes possess a sequence that is complementary to the mRNA start close to their 5' termini (TCCT at positions 15-18 which are complimentary to AGGA at positions 1-4; Barr *et al.*, 1991). The significance of this complementarity is unknown but it may relate to RNA secondary structure.

Gene	Gene start sequence		Gene end sequence	
	PVM	HRSV*	PVM	HRSV*
1C	AGGACAAGU	GGGGCAAU	UAGUUAUU	UAGUUAUU
1B	AGGAUAAU	GGGGCAAU	UAGUUAUAG	UAGUUAUU
N	AGGAUAAU	GGGGCAAU	UAUUUAUU	GAGUUAU
P	AGGACAAU	GGGGCAAU	UAUUUAUU	UAGUUAU
M	AGGAUAAU	GGGGCAAU	UAGUUAUU	AAGUUAU
SH	AGGAUAAGU	GGGGCAAU	UAGUUAU	UAGUUAUU
G	AGGACAAU	GGGGCAAU	UAGUUAU	UAGUUAUU
F	AGGAUGAGU	GGGGCAAU	UAGUUAUU	UAGUUAUU
22K	AGGACAAU§	GGGGCAAU	UAGUUAUU	UAGUUAUU
L	ND	GGGACAAA	ND	AUCUUAUU
consensus	AGGAUAAU	GGGGCAAU	UAGUUAUU	UAGUUAUU

Table 2: Gene start and end sequences for the 10 genes of HRSV and PVM.

The consensus sequences for these sequences are shown below the relevant columns. Upper case letters denote total conservation of a residue at that position in the sequence and lower case letters identify the most common residue at positions where conservation was not total. * The gene start and end sequences are for HRSV strain A2. §. The L gene start is located within the 22K gene. ND: not determined (adapted from Chambers *et al.*, 1990a).

1.5.4: Intergenic regions.

The intergenic regions of many paramyxoviruses such as SEN, PIV-3 and MV are amongst the most highly conserved regions both within each virus and also between the viruses and this high level of conservation implies that these sequences possess important roles. In contrast, the intergenic regions of PVM and HRSV show considerable variation both within each virus and between each other (Collins *et al.*, 1986; Chambers *et al.*, 1990b). This lack of homology in the pneumovirus intergenic regions may represent a difference in transcriptional function compared with the corresponding regions of many paramyxoviruses.

1.5.5: Pneumovirus RNA synthesis.

In contrast to other *Paramyxoviridae* family members, study of the processes of pneumovirus RNA synthesis has only recently begun and consequently these processes are not well understood at present. Herman (1989) developed a system capable of supporting transcription of the HRSV genome *in vitro* which employed a total extract of virus infected cells and was able to transcribe all the viral mRNAs. However attempts to transcribe purified nucleocapsids were not successful. Purified HRSV nucleocapsids were shown to be transcriptionally active after the addition of mock-infected cell extract (Barik, 1992) which suggested that a cellular factor was an obligate requirement of HRSV transcription. Inhibition of HRSV transcription using specific antibodies suggested that the cellular factor may be actin (Huang *et al.*, 1993).

An extremely powerful system has been developed for studying the genetics of HRSV which promises to greatly advance the understanding of processes involved in expression of the pneumovirus genome. Chimeric cDNAs were constructed which comprised (in 3' to 5' order) the HRSV leader sequence, the open reading frame of the chloramphenicol acetyltransferase (CAT) gene under the control of the HRSV gene

start and gene end signals, a one nucleotide long intergenic region, the complete HRSV L gene including the gene start and gene end signals and the 5' trailer region. The vRNA analogue encoded for by this cDNA was 7502 nucleotides long which in size represents almost half of the wild type HRSV genome (49.2%; Collins *et al.*, 1993). These cDNAs were transcribed *in vitro* and transfected into cells infected with wild-type HRSV. The synthetic vRNAs were amplified, expressed and packaged into virus particles by utilising the components supplied by the wild-type helper virus. The rescued particles were then used to infect fresh cells. A single nucleotide substitution in the 3' leader region of the vRNA analogue was shown to increase the efficiency of vRNA analogue rescue. By incorporating additional alterations of the vRNA analogue, this system allows characterisation of both cis-acting RNA sequences and trans-acting proteins involved in transcription, replication and packaging of the HRSV genome. This accomplishment paves the way for rescue of full-length synthetic HRSV genomes which will allow study of HRSV as a biological unit (Collins *et al.*, 1993). Hopefully, this system will allow the elucidation of the mechanisms involved in many aspects of pneumovirus genome expression and will consequently allow the structural differences between the pneumovirus and paramyxovirus genomes to be rationalised.

1.6: RNA SYNTHESIS IN NEGATIVE SENSE, SINGLE STRANDED RNA VIRUSES.

1.6.1: Introduction.

The ribonucleocapsid complex of the non-segmented negative strand RNA viruses is composed of the vRNA, and the L, P and N proteins. In these viruses, the L and P proteins have been identified as representing subunits of the viral RNA dependant RNA polymerase, and the N protein has been identified as the major structural component of the ribonucleocapsid complex which is responsible for encapsidation of the viral genome. In contrast to the N, P and L proteins of many members of the other major groups of non-segmented negative sense RNA viruses, the pneumovirus N, P and L proteins have not been fully characterised, however it is generally assumed that they perform analogous functions. Consequently, the overall mechanism of RNA synthesis in which the nucleocapsid components of these groups of viruses are involved is expected to be similar, if not entirely, at least in broad outline (Banerjee and Barik, 1992).

1.6.2: RNA synthesis in VSV.

Vesicular stomatitis virus has become the model non-segmented negative strand RNA virus due to the knowledge that has been amassed concerning its replicative and transcriptive processes. Emerson and Yu, (1975) initiated study of VSV RNA synthesis using a system that involved reconstituting purified nucleocapsid components *in vitro* to form a transcriptionally active ribonucleocapsid complex that was able to faithfully reproduce all the features of VSV transcription. Although this *in vitro* transcription system allowed many features of the N, P, and L proteins to be elucidated, the obligate requirement of both the L and P proteins for faithful transcription of mRNA species led to difficulties in assigning specific functions to each

of these protein species. A more recent and significant advance has been the opportunity to study the replication of VSV *in vivo* using a recombinant vaccinia virus T7 polymerase system capable of co-expressing all five genes of the VSV genome from cloned cDNA (Pattnaik and Wertz, 1990). By super-infecting cells with VSV defective interfering (DI) particles, this system is able to study the function each of these proteins performs in the process of VSV replication. The natural predisposition of DI particles for efficient replication means that study of the switch between transcription and replication is not possible using this system. The means with which to investigate VSV transcription was recently accomplished with the development of an *in vitro* reconstitution assay capable of faithfully transcribing purified nucleocapsid templates (Canter *et al.*, 1993).

1.6.3: Components of the VSV nucleocapsid.

(a) The VSV L protein.

The L protein has been suggested to encode all RNA transcriptional activities, namely, ribonucleotide polymerase (De and Banerjee, 1985), cap methylase (Hercyk *et al.*, 1988) and poly(A) polymerase (Hunt *et al.*, 1984). Less certain however is whether the L protein also encodes the protein kinase activity that phosphorylates the P protein. Highly purified L protein preparations made from VSV-infected cells contain an activity that specifically phosphorylates P protein and that is distinguishable from cellular kinases, thus suggesting that the activity came from the L protein itself and not a contaminating component (Barik and Banerjee, 1992). Work using a photoreactive analogue of ATP also suggested that a protein kinase activity resides on the native VSV L protein (Hammond *et al.*, 1992a). P protein phosphorylation was studied by Barik and Banerjee (1991) who used bacterially expressed and consequently unphosphorylated P protein as substrate for both cell-associated and L-associated protein kinase activity. The P protein was shown to exist in three different forms with

respect to their state of phosphorylation, namely the unphosphorylated P_0 , a partially phosphorylated P_1 and a more extensively phosphorylated P_2 . The P_0 form was converted to the P_1 form by a cell-associated kinase and P_1 was converted to P_2 by a purified L protein preparation. Studies by Chattopadhyay and Banerjee, (1987) revealed that the L protein kinase phosphorylated two serine residues located in the carboxyl-terminal region of the P protein, whereas the cell-associated kinase phosphorylated residues in the amino-terminus of the P protein. The substrate specificity exhibited by these two kinases necessitates that they act sequentially catalysing a cascade phosphorylation pathway. The L protein associated kinase appears to be a serine kinase (Hammond *et al.*, 1992b), although analysis of the L protein amino acid sequence fails to reveal any characteristic motifs indicative of serine or threonine kinase activity (Banerjee and Barik, 1992). It has been suggested that two P protein phosphorylating activities are present in the VSV virion, one of which resembles casein kinase II and is believed to be derived from the host cell (Beckes and Perrault, 1992).

(b) The VSV P protein

Reconstitution experiments have shown that whilst the L protein alone can initiate transcription on N-RNA templates to produce short oligonucleotide products, the P protein is required for the synthesis of authentic RNA products, suggesting that the P protein functions as a polymerase accessory factor or transcriptional activator (De and Banerjee, 1985). Additionally, *in vitro* studies have revealed that the P protein functions in maintaining the N protein in a replication competent state. This is accomplished by binding to and preventing N protein monomers from forming insoluble aggregates which are unable to support genome encapsidation (Howard and Wertz, 1989). It is in the form of the N-P complex that the N protein is believed to encapsidate nascent RNA molecules (Peluso and Moyer, 1988). N-P complexes able to

support *in vitro* genome encapsidation were only formed when the N and P proteins were concurrently synthesised. These complexes form in a variety of molar ratios (Masters and Banerjee, 1988), however a molar ratio of 1:1 was found to be essential for efficient genome replication (La Ferla and Peluso, 1989), and complexes of this ratio have been identified in VSV-infected cells (Moyer, 1989). The VSV P protein is one of the most highly phosphorylated proteins known, with over a sixth of the P protein amino acids representing potential phosphorylation sites. The phosphorylated amino acids are mainly concentrated in domain I, which is a negatively charged region spanning the amino-terminal third of the P protein (Bell and Prevec, 1985) and it has been suggested that this domain may act as a transcriptional activator of the L protein (Takacs *et al.*, 1991). Phosphorylation of residues in this region is believed to be mediated by the cellular kinase (Takacs *et al.*, 1991). The L protein of VSV is only able to bind to the N-RNA template in the presence of P protein (Mellon and Emerson, 1978). An L protein binding site has been located in domain I (Emerson and Schubert, 1987) and also between amino acids 213 and 247, designated domain II (Chattopadhyay and Banerjee, 1987). The carboxyl-terminal 21 amino acids of P is designated domain III and this region is required for binding to the N-RNA template (Emerson and Schubert, 1987). Phosphorylation studies have ascertained that serine is the major phosphate acceptor indicating that a serine kinase, a function that has been suggested for the L protein (Hammond *et al.*, 1992b), is responsible for phosphorylation of this protein. Reconstitution experiments have shown that only the fully phosphorylated form of P (P_2) is transcriptionally active, and that phosphorylation at serine residues 59 and 61 by a cellular kinase are essential for L protein phosphorylation that leads to functional activity of the P protein in the reconstituted transcription assay (Takacs *et al.*, 1992). This observation indicates that both the cell-associated and the L protein-associated kinases are essential for formation of functional P protein, a suggestion supported by restricted VSV replication in T-lymphocytes deficient in cellular protein kinase activity (Sleat *et al.*, 1992). The finding

that optimum transcriptional activity is only attainable with highly phosphorylated P protein is not in agreement with the findings of Beckes and Perrault (1992) who observed no loss in transcriptional activity when the P protein population was largely unphosphorylated. The phosphorylation status of the P protein was shown to regulate its ability to bind to the N-RNA template in the presence of L protein which is a crucial prerequisite for RNA synthesis (Chattopadhyay and Banerjee, 1987).

1.6.4: The model of RNA synthesis in negative sense, single stranded RNA viruses.

Kinetic studies have shown that the five VSV mRNA species are synthesised in the same order as their corresponding genes appear on the VSV genome and additionally that these mRNAs are transcribed in quantities that decrease as the distance of the corresponding gene from the 3' end of the genome is increased. The method by which the five VSV mRNAs are transcribed has not been elucidated, although two models are favoured (reviewed in Banerjee, 1987). Firstly, the stop-start model suggests that the polymerase binds to the genome at the 3' polymerase entry site and initially synthesises the leader sequence. The same enzyme then re-initiates RNA synthesis at the beginning of the N gene and continues until it meets the termination signal at which point the polymerase adds the poly(A) tail. The same enzyme either dissociates from the template or again re-initiates, this time at the beginning of the P (or NS) gene and so the process goes on. The important feature of this model is that once a polymerase molecule becomes removed from the genome, it may only re-attach to the genome at the 3' entry site. In this way, 3' proximal genes are abundantly represented by mRNA while 5' proximal genes are considerably less so. The other method, known as the multiple initiation model, supposes that rather than have a single 3' polymerase entry site, the polymerase may enter at several sites throughout the length of the genome. This model largely agrees with the stop-start method of

transcription in that mRNA synthesis results from separate initiation events, however it differs in that these initiation events are thought to be independent of one another.

Consistent with these models, initiation of both mRNA transcription and genome replication begins when the nucleocapsid complex binds to the 3' end of the genome and begins to synthesise the 47 base long leader sequence. The 3' terminal 15-17 nucleotides are required for optimal genome transcription and whilst certain changes are tolerated in this region, changes in the 3' terminal three nucleotides completely diminishes translational activity (Sherwood and Moyer, 1993). The finding that preservation of the initial 3' nucleotides is essential for transcription is consistent with the model that proposes a single 3' polymerase entry site. Up until this point, the events that have occurred are believed to be the same for both transcription and replication, and additionally, both the polymerase complex and also the RNA template are believed to be the same for both operations. The processes of transcription and replication have a fundamental difference in that genome replication requires concomitant protein synthesis whereas transcription does not (Wertz and Levine, 1973). The requirement for ongoing protein synthesis is met entirely by synthesis of the nucleocapsid protein (Patton *et al.*, 1984) and this led to the postulation of a mechanism for the switch between transcription and replication that depended upon the concentration of the N protein (Blumberg *et al.*, 1981). This mechanism relies upon the ability of the N protein to encapsidate the nascent RNA genome as it is synthesised. In this model, the polymerase complex begins to synthesise the positive polarity nascent leader RNA to which N protein molecules bind by recognition of an encapsidation signal (Blumberg *et al.*, 1983). The encapsidation signal is believed to reside within the first 10 nucleotides of the leader sequence, although for optimal encapsidation the first 19 nucleotides are needed (Moyer *et al.*, 1991). The encapsidation site is still active when nucleotide sequences are placed upstream of it, thus indicating that encapsidation by N protein molecules may proceed in both the 3' and 5' directions. Each N protein monomer is believed to associate with 6 nucleotides

(Blumberg *et al.*, 1983), a suggestion supported by Calain and Roux (1993) using SEN. As the polymerase complex passes down the genome, the growing RNA molecule is encapsidated by associating N protein molecules which somehow suppress a transcriptional termination signal which is thought to be located at the leader-N junction. The precise way in which the signal is suppressed is not known. The polymerase complex is thus able to continue RNA synthesis and eventually give rise to a full-length anti-genome molecule fully encapsidated in N protein molecules. If on the other hand, N protein was not present at an appropriate concentration, the leader RNA species would be synthesised but would not be encapsidated and consequently the transcription termination signal could not be suppressed, leading to the generation of the 47 base long leader. The polymerase would then either exit the genome or re-initiate transcription at the beginning of the N gene. This model suggests a self-regulating system where insufficient N protein stimulates increased mRNA synthesis to produce more viral proteins, and excess N protein triggers genome replication which will deplete the pool of viral proteins. Several observations have been made which give crucial support to this model such as the presence of tiny nucleocapsid structures containing the leader sequence alone (Blumberg *et al.*, 1981), and the presence of unencapsidated leader which represents the initial product of genome transcription. The existence of negative sense leader sequences also supports this model since it suggests that the polymerase complex initiating on anti-genomes (which are not templates for mRNA synthesis) is able to switch between a mode which makes full-length genomes to a mode in which genome synthesis is prematurely terminated at the postulated termination site, with the subsequent synthesis of a (-) leader. No provision has been made in this model for the role of the VSV matrix protein, which has been shown to inhibit transcription of the VSV genome both *in vitro* and *in vivo* (Kaptur *et al.*, 1991), although the recent existence of *in vivo* assays capable of studying viral transcription (Canter *et al.*, 1993) may soon allow the role of the M protein in transcription to be investigated.

1.6.5: The modified model of Paramyxoviridae RNA synthesis.

The level of understanding surrounding the mechanism of RNA synthesis in SEN is rapidly increasing and the recent development of *in vitro* assays which are capable of studying SEN replication (Calain and Roux, 1993) and transcription (Curran *et al.*, 1992) only serve to enhance this understanding. In some areas of study, the knowledge for SEN exceeds that which is known for any other negative sense, single stranded RNA virus. The information that is available for SEN suggest that whilst the replicative and transcriptive processes are similar to those of VSV, they do not precisely fit the model. For this reason a modified model for paramyxovirus transcription and replication has been proposed. Perhaps the most important difference between the RNA synthesis mechanisms of VSV and paramyxoviruses, such as MV and SEN, concerns the abundance of the leader sequence. In NDV, SEN and MV-infected cells, (+) leader sequences have only been identified in extremely low abundance, as have (-) leaders sequences (Kolakofsky *et al.*, 1991). In addition, two groups have identified large amounts of leader-N readthrough RNAs in MV-infected cells at an abundance 20-fold greater than that found for leader RNAs (Castadenda and Wong, 1989; Chan *et al.*, 1989). These results have been interpreted to suggest that the leader termination signal is either weak or absent in paramyxoviruses, and that the leader RNAs are not an essential operational element for paramyxovirus RNA synthesis (Blumberg *et al.*, 1991). Experiments using the Z and the H strains of SEN (both of which are considered to be wild type) have lead to the proposal that the polymerase responsible for transcription is distinct from the polymerase responsible for genome replication, since in the Z strain, they are distinguishable by the ability of the Z strain replicase, but not the Z strain transcriptase, to abort RNA synthesis at a point after the location of the leader-N junction (Vidal and Kolakofsky, 1989). The lesion responsible for the increased frequency of leader readthrough in the Z strain of SEN

has been mapped to the N protein gene, and not to the P or L genes, suggesting that the N protein plays a more active role in RNA synthesis besides that of a structural element (Curran *et al.*, 1993). It has also been proposed that whereas replication is believed to begin on the 3' end of the genome, transcription is believed to begin at an internally located site within the leader sequence, thus the paramyxovirus modified model employs an additional polymerase binding site to the VSV model.

The host cell factor tubulin has been shown to be an obligate requirement for transcription and replication in many negative sense, single stranded RNA viruses including measles, SEN and VSV (Moyer *et al.*, 1986, 1990). In contrast, actin is an obligate requirement for transcription of the HPIV-3 genome and the overall negative charge of the actin is thought to be the crucial factor affecting the transcriptional processes (De *et al.*, 1991). Using gel retardation techniques unidentified cellular factors have been shown to bind to the (+) strand leader and 3' non-coding region of the MV genome, although it is unknown if the association of these cellular factors with the genome was able to modulate the RNA synthetic capabilities of MV (Leopardi *et al.*, 1993). It has been suggested that cellular factors are also able to influence the mechanism that results in the production of edited P gene mRNA species (Horikami and Moyer, 1991). It is thus clear that cellular factors are able to influence viral mechanisms and because of this, the modified model does not suppose that all the processes of RNA synthesis are mediated exclusively by viral components.

The rhabdoviruses do not appear to be able to express proteins which are analogous to the V or C proteins of the paramyxoviruses, a situation that, given the ability of these proteins to interact with SEN nucleocapsid components (Yamada *et al.*, 1990) and influence RNA synthetic operations (section 1.4.3.1), would suggest possible subtle differences in the mode of RNA synthesis between these groups of viruses. The SEN C and C' proteins, which have been shown to be structural components of the SEN nucleocapsid (Yamada *et al.*, 1990), are capable of interacting

with nucleocapsid components and in particular to the P protein (Curran *et al.*, 1992). Whilst having no apparent affect on replication, the C proteins are capable of causing inhibition of transcription (Curran *et al.*, 1992). These workers speculated that the C proteins were, like N protein, able to act as transcriptional inhibitors, aiding the switch from transcription to replication. Two other products of the SEN P gene, namely the V and W proteins have been found to act as inhibitors of genome replication (Curran *et al.*, 1991a) and so perhaps only the P protein is essential for SEN RNA synthesis, the negative effectors being accessory factors which are dispensable. The suggestion that the role played by these proteins is non-vital is supported by the finding that several paramyxoviruses (including members of the SV5 and mumps group of paramyxoviruses, and also NDV) do not contain a C ORF, and PIV-1 which is very closely related to SEN (Lyn *et al.*, 1991) does not encode a V protein (Matsuoka *et al.*, 1991). It has been suggested that the V and C proteins may play an important role in establishing persistent MV-infections (Cattaneo *et al.*, 1989), and clearly any model for RNA synthesis will have to consider the role played by these species.

1.7: THE AIMS OF THIS STUDY.

The P genes of most *Paramyxoviridae* family members are able to direct the synthesis of at least one polypeptide, and in the case of the SEN P gene as many as eight (Curran *et al.*, 1991a). Previous studies have suggested that the PVM P gene is responsible for directing the synthesis of two polypeptides both *in vivo* (Ling and Pringle, 1989a), and *in vitro* (Chambers *et al.*, 1990a). The precise identity of these two polypeptides was not known but they were thought to either represent a single protein that underwent post-translational modification or the products of initiation on two in-frame P gene AUG codons. Confirmation that these two polypeptides were encoded by the PVM P gene would represent the first incidence of a pneumovirus being able to express more than one polypeptide from the P gene.

By determining the nucleotide sequence of a previously generated PVM P gene cDNA clone (Chambers *et al.*, 1990a) and by employing cell-free transcription and translation systems it was intended that the *in vitro* expressed polypeptide products of the PVM P gene could be characterised. The findings of the *in vitro* expression studies may then help elucidate the identity of the polypeptides expressed by the PVM P gene *in vivo*.

In contrast to the P gene, the N gene of *Paramyxoviridae* family members are believed to only to express one polypeptide product. Preliminary *in vitro* expression studies suggested that the PVM N gene was able to direct the synthesis of at least two polypeptides (Barr *et al.*, 1991). Further study of the *in vitro* expression characteristics of the N gene were undertaken to enable the precise identity of these species to be determined along with the mechanisms responsible for their synthesis.

An ability of the N and P proteins to interact has been demonstrated in many virus systems, although it has been particularly well studied in SEN (Homann *et al.*, 1991; Ryan and Kingsbury, 1988; Ryan and Portner, 1990; Ryan *et al.*, 1990; Ryan *et al.*, 1991). The development of an *in vitro* assay system to study the interaction

between the N and P proteins of PVM would identify the regions of each protein involved in the binding process.

CHAPTER 2

MATERIALS AND METHODS.

2.1: GROWTH OF PVM IN CELL CULTURE.

2.1.1: Growth of BSC1 cells.

BSC1 cells were cultured at an incubation temperature of 37°C in 1.5 litre glass roller bottles using GMEM (Glasgow modification of Eagles medium) growth medium supplemented with L-glutamine (4 mM), 5% [v/v] foetal calf serum, penicillin (1U/ml), and streptomycin (100 mg/ml), GMEM supplemented as such was referred to as complete GMEM. Cells were passaged by trypsinisation at 4 day intervals at a ratio of 1:4. The cells to be passaged were rinsed twice with 25 mls of versene, and then once with 25 mls of 20% (v/v) trypsin/versene solution. A further 10 mls of 20% trypsin/versene was added to the roller bottle which was rotated and agitated by hand until the cells became detached. The suspended cells were diluted in complete GMEM medium, and the appropriate amount of this suspension was introduced into a new roller bottle which was gassed with CO₂.

2.1.2: Growth of virus stocks.

The growth medium was poured off a roller bottle of sub- or just-confluent BSC1 cells and a small volume, typically 10-15 mls, of previously generated PVM stock was added to give a multiplicity of infection of between 0.1 and 0.5 plaque forming units (pfu) per cell. The bottle was incubated, with rolling, at 33°C for 2 hours after which 20 mls of GMEM (supplemented as above except for the addition of only 2% [v/v] foetal calf serum) was added. The PVM-infected BSC1 cells were incubated at 33°C until cytopathic effect (CPE) became extensive which typically took 4-6 days. The PVM-infected BSC1 cells were removed from the surface of the bottle by vigorous agitation and then stored at -70°C until needed.

2.1.3: Plaque assay titration of PVM stocks.

Virus stocks were serially diluted in complete GMEM to give a range of dilutions from 10^{-1} to 10^{-6} . A 1 ml volume of the diluted PVM stock was applied to confluent monolayers of BSC1 cells in 12 well tissue culture clusters (Costar). The monolayers were incubated for 2 hours at 33°C after which they were overlaid with complete GMEM containing 1% (w/v) Noble agar. When this had set, the clusters were incubated for 7 days at 33°C in a CO₂ gassed incubator. The cells were fixed for 3 hours by the addition of 1% (v/v) glutaraldehyde, after which the overlay was removed and the cells stained for 5 minutes with crystal violet (0.15% [w/v] in ethanol, diluted 1:20 before use). Excess stain was carefully rinsed away and the darkly coloured foci of PVM-infected BSC1 cells were counted using a low powered microscope.

2.1.4: Preparation of PVM-infected BSC1 cell material.

BSC1 cells were grown and infected with PVM stocks as described in section 2.1.2. After the CPE became extensive, the PVM-infected BSC1 cells were removed from the roller bottle using sterile glass beads and pelleted by centrifugation at 1000 g for 5 minutes at 4°C. The pelleted cells were resuspended in 1 ml of isotonic lysis buffer (150 mM NaCl, 20 mM Tris-HCl [pH 7.5], 2 mM EDTA, 1% Nonidet P-40 [BDH laboratory supplies]) and incubated on ice for 15 minutes. The suspended cell material was centrifuged at 5000 g for 5 minutes. The supernatant fraction which contained the solubilised PVM-infected BSC1 cell proteins was retained and divided into 100 µl aliquots which were stored at -20°C.

2.2: PROTEIN ANALYSIS.**2.2.1: Polyacrylamide gel electrophoresis.**

The discontinuous buffer system of polyacrylamide gel electrophoresis first described by Laemmli (1970) was utilised. In all the polyacrylamide gel electrophoresis analyses described in this thesis, a single concentration 10% (w/v) resolving gel was used. The composition of the resolving and stacking gels used for electrophoresis is shown below.

Reagent	volume in resolving gel	volume in stacking gel
30% (w/v) acrylamide:bisacrylamide (29:1)	16.7 ml	4 ml
Tris-HCl (3 M, pH 8.8)	6.7 ml	----
Tris-HCl (0.5 M, pH 6.8)	----	5 ml
10% (w/v) sodium lauryl sulphate	0.5 ml	0.2 ml
10% (w/v) ammonium persulphate	0.5 ml	0.1 ml
TEMED	0.006 ml	0.003 ml
distilled water	25.6 ml	10.7 ml

The resolving gel acrylamide mixture was poured between glass plates and overlaid with water saturated butanol and allowed to polymerise. After the gel had set, the butanol was flushed off and the stacking gel acrylamide mixture was added, into which a comb was inserted to form the sample loading wells. Samples were prepared for electrophoresis by the addition of the appropriate volume of 5x SDS-PAGE sample buffer (15% [v/v] β -mercaptoethanol, 15% [w/v] SDS, 50% [w/v], glycerol,

bromophenol blue), and then boiled for 3 minutes. The samples were loaded onto the gel which was electrophoresed in SDS running buffer (192 mM glycine, 25 mM Tris, 50 mM SDS, pH 8.3) using a Protean® II xi gel apparatus (Biorad) for a time period and current rating that caused sufficient separation of the polypeptides under examination. This was typically 14 hours at a current of 12 mA. Depending on the nature of the analysis, the electrophoresed gels could then be processed in one of three ways. Firstly, gels could be stained using Coomassie blue R-250 (0.1% [w/v] in 16% [v/v] acetic acid, 42% [v/v] distilled water, 42% [v/v] methanol) to directly visualise the separated polypeptides. Secondly, gels on which radioactively labelled polypeptides were electrophoresed were fixed for 45 minutes in a solution composed of 10% (v/v) acetic acid 45% (v/v) methanol and 45% distilled water, dried and exposed to X-ray film (Fuji photographic film Co. Ltd.). Or thirdly, the polypeptides separated on the gel were transferred to a PVDF membrane for Western or west-western blotting, described in section 2.2.2.

2.2.2: Transfer of proteins from polyacrylamide gel to PVDF membrane.

Following electrophoresis, the polyacrylamide gel was placed in 500 mls of transfer buffer (192 mM glycine, 25 mM Tris, 20% [v/v] methanol, pH 8.3) and gently agitated for 30 minutes at room temperature. The gel was then placed in contact with a similar sized piece of Immobilon™ PVDF membrane (Millipore Corporation) which had been prepared according to the manufacturers recommendations. All bubbles and air pockets that had become trapped between the gel and membrane were carefully removed and the gel/membrane sandwich was positioned in a Biorad Transblot™ apparatus according to instructions supplied by the manufacturer. Proteins were transferred by electrophoresis for 3 hours at 70 V. In order to visualise the immobilised proteins, the membrane was incubated for 5 minutes in the general protein stain ponceau-S (0.5% Ponceau S [w/v] in 1% [v/v] acetic acid 99% [v/v] distilled water)

and then destained in distilled water. The position of lanes and molecular weight standards were lightly marked using a pencil. Any residual stain was removed by further washing in distilled water. At this point, the membrane could either be used for Western blot analysis, west-western blot analysis or could be stored at -20°C for future use

2.2.3: Western blot analysis.

Proteins separated by SDS-PAGE were transferred to Immobilon™ PVDF membrane as described above. In order to minimise non-specific binding of antibody to the membrane matrix, the membrane was incubated for 20 minutes at room temperature in Tris-buffered saline (TBS, 100 mM Tris, 150 mM NaCl, [pH 7.6]) containing 5% (w/v) Marvel™ dried milk powder (Premier brands Ltd.). The membrane was thoroughly washed in TBS containing 0.1% (v/v) poly-oxyethylene-sorbitan-monolaurate (Tween-20, Sigma Chemical Co.) and then incubated with the appropriate antibody at an optimal dilution that had been previously determined. Unless otherwise indicated, membrane washes consisted of 3 washes for 1 minute followed by four washes for 4 minutes. The incubations involving antibodies were usually carried out in heat-sealed plastic bags which allowed volumes to be kept to a minimum. The membrane was incubated with the antibody for periods of time ranging from 1 hour to 16 hours. The 16 hour antibody incubations were performed at 4°C. The membrane was washed as described above and incubated for between 20 minutes and 1 hour at room temperature with a biotinylated second antibody, which was diluted in TBS by a previously determined factor which gave the optimal signal above background. The membrane was washed as before, and incubated for 20 minutes with an alkaline phosphatase linked streptavidin conjugate (Amersham International plc.) diluted 1:3000 in TBS. After washing, the membrane was incubated with alkaline phosphatase substrates (Nitro-blue tetrazolium and 5-bromo-4-chloro-3-indolyl

phosphate). As soon as the required intensity of the appropriate bands was reached, the reaction was stopped by the addition of distilled water.

2.2.4: The west-western blot assay.

The west western protein binding procedure described here is based on the method described in Homann *et al.*, (1991). The procedure described here was able to study the interactions that occurred between PVM protein species which were immobilised onto a membrane support and radiolabelled PVM nucleocapsid protein that was applied to the membrane. The assay relied upon the gradual removal of detergent to hopefully allow the immobilised protein species to regain their functional conformation.

Proteins separated by SDS-PAGE were transferred to PVDF membrane in a method identical to that used in the transfer of proteins for Western blot analysis except for the exclusion of methanol from the transfer buffer. Proteins immobilised on the membrane were renatured for 12 hours in standard binding buffer (SBB, 10 mM HEPES [*N*-2-hydroxyethylpiperazine-*N'*-2-ethane-sulphonic acid] [pH 7.4], 10 mM MgCl₂, 50 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, 10% glycerol). Following this incubation, the membrane was incubated for 3 hours at room temperature in SBB containing 5% (w/v) dried milk powder in order to saturate free binding sites present on the membrane matrix. The blot was rinsed in SBB and cut into strips, each of which had a PVM-infected BSC1 cell track and a mock-infected BSC1 cell track. Each strip was incubated for 24 hours at 4°C in 1 ml of SBB containing 30,000 cpm of trichloroacetic acid precipitable activity of *in vitro* translated N protein or N protein related product. Following this incubation, the strips were each washed 10 times for 10 minutes in SBB, dried in a 70°C oven, and exposed to pre-flashed X-ray film overnight. The extent of binding that had occurred was assessed by densitometric analysis.

2.2.5: Densitometric analysis.

Pre-flashed autoradiographs generated by the west western blotting procedure were analysed using Image-Quant (version 3.0) software running on a Molecular Dynamics computing densitometer.

2.3: MANIPULATIONS IN BACTERIA.

2.3.1: Bacterial Strains.

(a) *Escherichia coli* TG2.

Escherichia coli K12 strain TG2 was used for all general cloning procedures. The TG2 genotype is supE hsd^Δ5 thi ^Δ(lac-proAB) ^Δ(srl-recA)306::Tn10(tet^r) F'[traD36 proAB⁺ lacI^q lacZ^ΔM15] (Sambrook *et al.*, 1989). *E. coli* strain TG2 was maintained on 2YT agar plates (16 g bacto-tryptone, 10 g bacto-yeast extract, 5 g NaCl, 1.5% [w/v] agar, in 1 litre distilled water) which were stored at 4°C. To initiate cultures of *E. coli* TG2, a single colony of bacterial cells was picked from a maintenance plate using a sterile toothpick, and used to inoculate a 5 ml volume of 2YT media (as above, without agar) contained in a sterile glass universal bottle. Cultures were incubated overnight at 37°C in an orbital shaker.

(b) *Escherichia coli* BL21.

E. coli strain BL21 was used for the *in vitro* expression of PVM P protein and P protein related species. This bacterial strain contains a chromosomal copy of the T7 RNA polymerase gene under the control of the isopropyl- β -D-thiogalactoside (IPTG) inducible *lac* promoter (Studier and Moffat, 1986). The genotype is hsdS gal (λ clts857 ind1Sam7 nin5 lacUV5 T7 gene1) *E. coli* strain BL21 was maintained on 2YT agar plates which were kept at 4°C.

(c) *Escherichia coli* M15 [pREP4].

E. coli K12 strain M15 [pREP4] was used for the high level expression of recombinant proteins using the pQE series of plasmids (Qiagen Inc.). The genotype of the strain is NaI^s , Str^s , rif^s , lac^- , ara^- , gal^- , mtl^- , F^- , recA^+ , uvr^+ . *E. coli* strain M15 included multiple copies of the plasmid pREP4 which contains the *lac* I gene ensuring high levels of lac repressor and tight regulation of protein expression.

2.4: ROUTINE MANIPULATIONS OF DNA.**2.4.1: Phenol and chloroform extraction.**

Phenol extractions were performed to remove contaminating protein from samples of nucleic acid. An equal volume of phenol (BDH lab. supplies) equilibrated with TNE (100 mM NaCl, 50 mM Tris-HCl [pH 7.5], 1 mM EDTA) was added to the sample to be extracted which was then vortexed for 30 seconds. The sample was centrifuged to separate the phases and the upper aqueous layer was transferred to another tube to which an equal volume of chloroform was added. The sample was

vortexed and centrifuged and again the upper aqueous layer taken for further manipulation.

2.4.2: DNA and RNA precipitations.

Nucleic acid containing solutions were made 0.3 M with respect to sodium acetate (pH 5), and 2.5 volumes of absolute ethanol were added. Nucleic acid was precipitated by incubating the samples at -20°C overnight or in a dry-ice and ethanol bath for 3-5 minutes. Nucleic acid was pelleted by centrifugation.

2.5: BACTERIAL TRANSFORMATION AND PREPARATION OF PLASMID DNA.

2.5.1: Preparation of competent bacteria.

A 1 ml aliquot of an overnight culture of *E. coli* TG2 was used to inoculate 70 mls of 2YT medium in a sterile conical flask. The resulting culture was incubated with shaking in a 37°C incubator. When the optical density (measured at 590 nm) of the culture reached 0.4, the culture was removed from the incubator and placed on ice for 30 minutes. The cooled culture was centrifuged at 2,500 g for 5 minutes to pellet the bacteria. The bacteria were resuspended in 10 mls of sterile, ice cold, 0.1 M MgCl₂ and then immediately repelleted by further centrifugation at 2,500g for 5 minutes. The pelleted bacteria were resuspended in 2 mls of sterile, ice cold, CaCl₂ and placed on ice. After 1 hour the *E. coli* TG2 were sufficiently competent to be used for plasmid transformation.

2.5.2: Transformation of *E. coli* TG2 with plasmid DNA.

Aliquots (100 µl) of competent *E. coli* TG2 were dispensed into ice cold sterile Falcon tubes (Beckton-Dickenson labware). Plasmid DNA generated as a result of ligation reactions, typically 25 µl in volume, was added to the *E. coli* TG2 cells and the tubes were returned to ice for 30 minutes. The *E. coli* TG2 were heat shocked by placing the tubes in a 42°C water bath for 2 minutes, and then returned to ice for between 1 and 5 minutes. 2YT medium (0.5 ml) was added to each tube which was then placed in a 37°C shaking incubator for 20 minutes. The resulting cultures were spread onto 2YT plates supplemented with a selective antibiotic which was most commonly ampicillin (100 µg/ml final concentration). Once they had dried, plates were placed in a 37°C incubator overnight.

2.5.3: M13 DNA transformation.

Transformation of M13 DNA into *E. coli* TG2 was performed using a method essentially the same as that described for transformation of *E. coli* TG2 with plasmid DNA, until after the heat shock stage. At this point, M13 transformed *E. coli* TG2 were chilled on ice for 5 minutes and then warmed to room temperature. A 2 ml aliquot of top agar (2YT containing 0.6% [w/v] agar) at a temperature of 45°C was then added to the transformed bacteria. The top agar contained both X-gal (3 mg/ml) and IPTG (2.4 mg/ml) for chromogenic identification of recombinants, and also 100 µl of early log phase *E. coli* TG2 to provide a bacterial lawn. The Falcon tube was briefly agitated and the contents poured on a warm 2YT agar plate which was left to set on a level surface. Plates were incubated overnight at 37°C.

2.5.4: Small scale plasmid preparations.

Plasmid DNA was extracted from bacterial cultures using a modified version of the alkaline lysis method described by Sambrook *et al.*, (1989). Transformed colonies were picked from agar plates using sterile toothpicks and inoculated into 2 mls of 2YT medium supplemented with the appropriate antibiotic. Cultures were incubated at 37°C overnight in a shaking incubator. 1.5 ml of each culture was centrifuged at 10,000 g for 2 minutes in an microfuge (Eppendorf Geratebau Netheler+Hinz GmbH) to pellet the bacteria, and the supernatant was discarded. The resulting pellet was resuspended in 100 µl of lysozyme solution (10 mg/ml in distilled water) and placed on ice for 10 minutes. 200 µl of SDS-lysis buffer (0.2 M NaOH, 1% [w/v] SDS) was added to the cells and the resulting suspension was mixed by gently agitating and flicking the tubes. The tubes were incubated on ice for 20 minutes, after which 150 µl of 3 M sodium acetate (pH 5) was added. The resulting viscous material was mixed by gentle agitation as before. The precipitate of cell debris was pelleted by centrifugation at 10,000 g for 5 minutes and the supernatant fraction was transferred to a fresh Eppendorf tube containing 1 ml of ethanol. DNA was precipitated by incubating the tube in dry-ice/ethanol for 3 minutes and the DNA was pelleted by centrifugation at 10,000 g in a microcentrifuge. The DNA pellet was dried in a vacuum desiccator, resuspended in 100 µl of distilled water and stored at -20°C. This basic method of plasmid preparation was occasionally modified to accommodate larger cultures up to 60 mls in volume by scaling up the quantities of reagents such that their proportions to one another remained the same as above.

2.5.5: Large scale plasmid preparations.

The alkaline lysis method of Sambrook *et al.* (1989) was used. An overnight culture of the appropriate transformed *E. coli* TG2 colony was used to inoculate 500 mls of 2YT medium containing the appropriate antibiotic. The culture was incubated in

an orbital shaker at 37°C for 6 hours. After this period, chloramphenicol (20 µg/ml, final concentration) was added in order to amplify plasmid numbers (Sambrook *et al.*, 1989). The culture was incubated overnight and the bacteria were harvested by centrifugation at 2,000 g for 30 minutes. The resulting pellet was resuspended in 7 mls of solution I (25 mM Tris-HCl [pH 8.0], 50 mM glucose, 10 mM EDTA, 5 mg/ml lysozyme), transferred to an Oakridge tube and then incubated for 5 minutes at room temperature. The bacteria were lysed by the addition of 14 mls of solution II (0.2 M NaOH, 1% [w/v] SDS) and the tube was gently agitated to ensure thorough mixing of the contents. After a 10 minute incubation on ice, 10 mls of solution III (5 M sodium acetate, [pH 5.0]) was added and once again, the tube contents were gently mixed and then incubated on ice for 10 minutes. The mixture was centrifuged at 5,000 g for 30 minutes to pellet cell debris and chromosomal DNA. The supernatant was mixed with 0.6 volumes of isopropanol, and after a 10 minute incubation at room temperature, the DNA was precipitated by centrifugation at 2,500 g for 20 minutes. The DNA pellet was dried in a desiccator and resuspended in 25 mls of distilled water to which 26.5 g of caesium chloride and 3 mg of ethidium bromide were added. The solution was introduced into a Beckman 38 ml quick seal™ tube (Beckman Instruments Inc.), and the tube was filled with paraffin. The tube was heat sealed and centrifuged for 16 hours at 60,000 g in a Beckman Vti50 rotor. The closed circular plasmid DNA band was located using long wave (366 nm) UV light and the band was removed by piercing the tube with a syringe. The ethidium bromide was removed by four extractions using an equal volume of water saturated butanol. The CsCl was removed by dialysis against three 3 litre volumes of TE (1 mM Tris-HCl [pH 7.8], 1 mM EDTA [pH 8.0]). The DNA was recovered by ethanol precipitation and resuspended in 1 ml of distilled water. The yield was calculated by measuring the OD₂₆₀ (an OD₂₆₀ of 1.0 equates to 50 µg/ml of ds DNA).

2.5.6: Single strand template preparation.

Turbid plaques were transferred into 2 mls of 2YT medium which contained 60 μ l of an overnight *E. coli* TG2 culture. To achieve good aeration, glass test tubes with loose fitting aluminium tops were used (Difco Laboratories). Cultures were shaken at 200 rpm. for between 4 and 6 hours. After incubation, the culture medium was decanted into a 1.5 ml Eppendorf tube and the bacteria were pelleted in an Eppendorf microfuge. Approximately 1 ml of supernatant was transferred to a clean Eppendorf tube to which 250 μ l of a solution of PEG 6000 (20% [w/v]) and NaCl (2.5 M) was added and mixed. The tube was incubated at room temperature for 15 minutes after which M13 bacteriophage were pelleted by centrifugation at 10,000 g for 5 minutes. The pellet was resuspended in 120 μ l of TE buffer (10 mM Tris-HCl [pH 8.0], 0.1 mM EDTA) and extracted with 50 μ l of TNE equilibrated phenol. A 100 μ l volume of the aqueous phase was removed and extracted with 50 μ l of chloroform. Following chloroform extraction, 80 μ l of the aqueous phase was removed and the DNA precipitated as described previously. The DNA pellet was resuspended in 30 μ l of sterile distilled water.

2.6: MODIFICATION OF DNA.

2.6.1: Restriction enzyme digestion of DNA.

Double stranded DNA was specifically cleaved using various restriction endonucleases obtained from commercial sources. Optimal cleavage of DNA was achieved by performing the digestion under the conditions recommended by the manufacturer. Typically restriction enzyme digestions were carried out in a 10 μ l volume, comprising approximately 0.1 μ g of plasmid DNA, sufficient restriction

enzyme to ensure complete cleavage of the DNA and the correct concentration of the appropriate buffer. If the plasmid DNA was obtained using the small scale preparative method then RNase A (Sigma Chemical Co.) was also included in the restriction reaction at a final concentration of 100 µg/ml.

2.6.2: End-filling of DNA fragments.

Ends of DNA fragments were made blunt either by nibbling back a 3' protruding strand or by filling in a 5' protruding strand to make it flush. Both manipulations were performed using DNA polymerase I Klenow fragment (Bethesda research laboratories) and 100 µm dNTPs (Pharmacia LKB), in 1X React™ 2 enzyme buffer (Bethesda research laboratories). The reactions were incubated for 30 minutes at 37°C

2.6.3: Ligations.

Fragments of DNA generated by the fragment preparation method described below were inserted into linearised plasmid vectors by ligation using T4 DNA ligase (Amersham international plc.). Typical ligation reactions were carried out in a volume of 50 µl which comprised linearised plasmid vector, purified DNA fragment, a sufficient quantity of T4 DNA ligase and an appropriate quantity of T4 DNA ligase buffer (Bethesda research laboratories). Reactions were performed at room temperature for between 3 and 16 hours.

2.7: ELECTROPHORESIS OF DNA.

2.7.1: Agarose gel electrophoresis.

This technique was used analytically to resolve the products of restriction enzyme digestions and also preparatively to isolate specific DNA fragments. Agarose (Sigma Chemical Co.) was dissolved in near boiling TAE buffer (40 mM Tris-acetate [pH 7.0], 1 mM EDTA). The molten agarose was cooled and ethidium bromide was added to give a final concentration of 0.5 µg/ml. The agarose concentration within the gel varied between 0.6% and 2.0% depending on the size of DNA fragments that had to be resolved. The relationship between agarose concentration and the resolving characteristics of a gel are detailed in Sambrook *et al.* (1989). The cooled agarose was poured into a gel casting assembly and combs were inserted in order to create sample wells. DNA samples to be analysed were mixed with loading buffer and applied to the sample wells. In order to determine the molecular weight of resolved bands, 1 kb ladder markers (Bethesda Research Laboratories) were loaded alongside the samples under analysis. Set agarose gels were electrophoresed in one third strength TAE buffer which also contained ethidium bromide (0.5 µg/ml). Gels were usually electrophoresed at a power rating of 45 W for a period of time which allowed sufficient separation of the DNA fragments. Resolved DNA bands were visualised by placing the gel on a UV transilluminator (UVP Inc.). Photographs were taken using Polaroid 667 instant black and white film in a Polaroid DS-34 camera.

2.7.2: Polyacrylamide gel electrophoresis.

Acrylamide and bis-acrylamide in a ratio of 29:1 were dissolved in a solution of TBE (89 mM Tris-HCl [pH 7.5], 89 mM Boric acid, 10 mM EDTA) and 8 M urea to give an overall acrylamide concentration of 6%. To ensure rapid polymerisation of

the gel, 120 µl of TEMED and 600 µl of 10% (w/v) ammonium persulphate were added. The gel was poured between clean glass plates separated by spacers designed to give a wedge shaped gel 0.4 mm thick at the top and 1.0 mm thick at the bottom. Shark tooth gel combs (Gibco-BRL) were placed at the top of the gel to form a straight edge on which the samples could later be applied, and the gel was left to polymerise for 1 hour. Before samples were loaded, the gel was pre-electrophoresed for 15 minutes at a constant power rating of 80 W. Samples were prepared as described in the protocol supplied with the Sequenase™ version 2.0 kit (United States Biochemical Corp., see section 2.9) and then were electrophoresed for periods of time of between 1.5 hours and 6 hours depending upon the position of the required sequence in relation to the -20 primer. When electrophoresis was complete, plates were separated and the gel was fixed in a solution of 10% acetic acid (v/v). The gel was dried on a vacuum drier (Biorad) and then exposed to x-ray film (Fuji photographic Film Co. Ltd.).

2.7.3: Fragment preparations.

Restriction endonuclease generated DNA fragments were recovered from agarose gels by a method based on the freeze-squeeze procedure described by Thuring *et al.* (1975). An agarose slice containing the appropriate DNA band was cut from the gel, chopped into small pieces, and then placed in a 1.5 ml Eppendorf tube. DNA was extracted from the agarose by repeated freezing and thawing of the tube contents which released DNA into the expelled liquid. Once extraction of liquid from the agarose became difficult the gel pieces were wetted with a small quantity of distilled water (100 µl) and subjected to further freeze-thaw cycles. The resulting liquid was extracted three times with phenol and once with chloroform, as described in section 2.4.1 after which the DNA was precipitated as described in section 2.4.2.

2.7.4: Polymerase chain reaction.

The polymerase chain reaction (PCR) amplification technique used in this study was based on a method described by Saiki *et al.* (1989). Elongation of the DNA strands was performed by *Taq* I polymerase (Promega Corporation) in reaction conditions stipulated by the manufacturer. A typical PCR amplification was performed in a 100 µl volume consisting of the following reagents.

<i>Taq</i> I polymerase (Promega Corporation)	0.2 µl
100 mM MgCl ₂ (Promega Corporation)	10.0 µl
10x <i>Taq</i> I buffer (Promega Corporation)	6.0 µl
dNTP mix (dGTP, dCTP, dTTP, dATP, all at 25 mM)	1.0 µl
oligonucleotide primers (approximately 40 µg/ml)	1.0 µl
DNA template (1 ng/ml)	2.0 µl
distilled water	80.0 µl

The PCR reaction components were overlaid with 100 µl of white paraffin to limit volume changes due to evaporation. The samples were placed in a thermal reactor (Hybaid) and subjected to a 30 cycle programme which involved a 45 second melting step (94°C), a 45 second annealing step (55°C) and a 90 second elongation step (74°C)

2.8: CHARACTERISATION OF PVM GENE 4 CDNA CLONES.

A library of double stranded cDNA clones generated from mRNA extracted from PVM-infected BSC1 cells was constructed by Dr. P. Chambers (Chambers *et al.*, 1990a). A group of cDNA clones was identified which had a high relative abundance

of members and which hybridised with virus specific RNA with an approximate molecular weight of 1200 bases. This group was assigned to the PVM N gene. Nucleotide sequence analysis of cDNA clones from the group confirmed this assignment (Barr *et al.*, 1991). By nucleotide sequence analysis of polycistronic clones, the group that represented the proceeding gene to the N gene was identified. One of the cDNA clones from this group (clone 4.48) represented the full length of PVM gene 4 although this clone contained a sequence inversion at its 5' end (Barr *et al.*, 1991).

A population of oligo(dG) tailed cDNA-RNA hybrid molecules generated from PVM-infected BSC1 cell extracted RNA (Barr *et al.*, 1991) were used as a template for the amplification of a cDNA segment corresponding to nucleotides 1 to 139 of the PVM P gene using PCR. One of the PCR oligonucleotide primers (designated PVMP1, GGTGACATGAGTAGTGGCAGTGTTCGG) was designed to hybridise to a region of the P gene spanning nucleotides 115 to 139, whereas the other primer (designated oli C, CCCGAGCTCTGCAGGATCCCCCCCCCCCCC) was designed to hybridise to the poly(dG) tract. The DNA segment was amplified and purified as described in Barr *et al.*, (1991). The inverted 5' sequence of cDNA clone 4.48 was removed by digestion with restriction enzymes *Acc* I and *Pst* I and this sequence was replaced with the PCR generated segment, previously cleaved with *Acc* I and *Pst* I. This plasmid, containing a full-length cDNA copy of the PVM P gene, was designated pUC13-P.

2.9: NUCLEOTIDE SEQUENCE ANALYSIS.

Nucleotide sequence was determined using the principle of dideoxy chain termination developed by Sanger *et al.* (1980). The protocol and reagents excepting radioactive label were supplied in the Sequenase™ kit version 2.0 (United States Biochemical Corp.) The ³⁵S-dATP radioactive label was supplied by Amersham International plc. The -20 sequencing primer supplied in the Sequenase™ kit was used for all sequence determinations.

2.10: IDENTIFICATION OF ANTI-PVM P PROTEIN SPECIFIC MONOCLONAL ANTIBODIES.

Approximately two hundred monoclonal antibody secreting hybridoma cell lines were generated by Dr. R. Ling (Ling and Pringle, 1989a). Tissue culture fluids from each of the cell lines, collected and stored at -20°C by Dr. Ling, were available and were screened either for the ability to recognise bacterially expressed PVM P protein by Western blot analysis or for the ability to immunoprecipitate ³⁵S methionine labelled, *in vitro* translated, P protein. In this way, several cell lines capable of secreting anti-P protein monoclonal antibodies were identified, and for one of these cell lines, 26/3/B5, the original liquid nitrogen-stored cell culture was resurrected. Frozen cells were quickly thawed and transferred to a sterile universal containing 5 mls of RPMI 1640 medium (Gibco-BRL) supplemented with foetal calf serum (10% v/v), penicillin (100 U/ml), streptomycin (100 µg/ml), glutamine (2 mM), HEPES (40 mM), and sodium hydrogen carbonate (0.2% w/v). Medium containing these additions was referred to as complete. The tube was immediately centrifuged at 500 g for 2 minutes

to pellet the hybridoma cells. The cell pellet was gently resuspended in 100 µl of complete RPMI 1640 medium and dispensed onto macrophage feeder cell cultures, in a 96 well tissue culture dish. The macrophage feeder cells were obtained by rinsing the peritoneal cavity of a freshly killed mouse with 5 mls of PBS, withdrawing the PBS with a syringe, and centrifuging the resulting cell suspension at 200 g for 5 minutes. The feeder cells were resuspended in complete RPMI 1640 medium and cultures were initiated by dispensing aliquots of the cell suspension into a 96 well dish (Costar). The hybridoma cells were incubated at 37°C in a 5% CO₂ atmosphere and after 4 days were transferred, using a Pasteur pipette, for culture into 12 well dishes without the feeder cells. As cell numbers increased the cells were transferred to progressively larger culture flasks using a cell splitting procedure identical to that used for BSC1 cells described in section 2.1.1. Cell culture medium was removed each time the cells were split and used in Western blot analysis.

2.11: EXPRESSION OF RADIOLABELLED POLYPEPTIDES *IN VITRO*.

2.11.1: Transcription of mRNA *in vitro*.

mRNA was transcribed *in vitro* from derivatives of the T7 transcription plasmid pGEM1 (Promega Corp.). Transcription plasmids were linearised by digestion with an appropriate restriction enzyme that cut at a site downstream from the termination codon of the coding region present on each plasmid. Messenger RNA was transcribed from the T7 promoter using T7 RNA polymerase under reaction conditions recommended by the enzyme manufacturer (BRL). Typically, transcription reactions were performed in 50 µl volumes consisting of the following reagents.

linearised plasmid (approx. 0.1 mg/ml)	10.0 µl
5x T7 buffer (supplied by BRL)	10.0 µl
ATP, UTP, CTP, GTP combined stock, each at 25 mM	2.0 µl
T7 RNA polymerase (10 U/µl)	1.0 µl
distilled water	27.0 µl

The transcription reactions were allowed to proceed at 37°C for 1 hour after which a further 5 units of T7 RNA polymerase was added. After incubation for a further 30 minutes, the volume of the transcription reactions was increased to 200 µl by the addition of distilled water. The transcription products were extracted once with phenol and once with chloroform, (as described in section 2.4.1) and the mRNA was precipitated by the addition of 500 µl of ethanol and incubation for 30 minutes at -70°C. The RNA was harvested by centrifugation at 10,000 g for 10 minutes and the pellet was dried and resuspended in 50 µl of distilled water.

2.11.2: *In vitro* translation using rabbit reticulocyte lysate.

Messenger RNA species transcribed from the pGEM1 derived plasmids described above, were used in conjunction with a rabbit reticulocyte lysate system (Amersham International plc.) to programme the *in vitro* translation of radioactively labelled polypeptides. The constituents of a typical *in vitro* translation reaction are shown below.

rabbit reticulocyte lysate	10.0 μ l
mRNA	1.0 μ l
35 S methionine	1.0 μ l
distilled water	0.5 μ l

If larger amounts of *in vitro* synthesised polypeptides were required, these constituents were scaled-up such that they remained in these proportions. The translation was allowed to proceed for 90 minutes at 30°C. The incorporation of 35 S methionine into polypeptide translation products was assessed either by subjecting an aliquot of the lysate to SDS-PAGE analysis, or by trichloroacetic acid precipitation. The concentration of mRNA required in the translation reaction to result in optimal synthesis of the desired polypeptide was determined each time a new batch of mRNA was transcribed by serially diluting the mRNA added to the reticulocyte lysate and analysing the resulting products with SDS-PAGE.

2.11.3: *In vitro* translation using wheat germ lysate.

In addition to rabbit reticulocyte lysate, a wheat germ lysate was sometimes used for the *in vitro* expression of 35 S methionine labelled polypeptides. The wheat germ lysate and the energy mix were kind gifts of Dr. Martin Hartley, University of Warwick, UK. The constituents of a typical wheat germ lysate translation reaction, and composition of the energy mix is shown below.

(a) Wheat germ translation constituents.

Wheat germ lysate	3.8 μ l
energy mix	2.4 μ l
³⁵ S methionine	1.0 μ l
distilled water	3.4 μ l
mRNA	1.0 μ l

Translations were allowed to proceed for 1 hour at 27°C.

(b) Energy mix composition.

HEPES (1 M [pH 7.6])	70 μ l
ATP (0.1 M)	50 μ l
creatine phosphate (0.4 M)	100 μ l
creatine phosphokinase	20 μ l
spermidine (pH 7.0)	100 μ l
DTT (0.1 M)	85 μ l
GTP (2 mM)	50 μ l
amino acid mix (all at 1 mM)	25 μ l
potassium acetate (1.0 M)	20 μ l
distilled water	420 μ l

2.11.4: Immunoprecipitation of ^{35}S methionine labelled polypeptides.

Radiolabelled polypeptides were produced by *in vitro* translation using the rabbit reticulocyte system described earlier. The translation mix was diluted by the addition of 6 volumes of IVT precipitation buffer (50 mM Tris-HCl [pH 7.6], 5 mM EDTA [pH 8.0], 150 mM NaCl, 0.05% [w/v] Nonidet P-40) in a 1.5 ml Eppendorf tube and chilled on ice. Antibody originating from either hyper-immune animal serum or hybridoma cell culture fluids was added to the diluted translation product and the mixture was incubated overnight at 4°C. The precise volume of the antibody solution that was added depended on the titre. After the overnight incubation, 100 µl of Sepharose CL-4B linked *staphylococcus* protein A slurry (Pharmacia LKB), prepared according to the manufacturers instructions, was added to the immunoprecipitation mixture. The resulting suspension was vigorously agitated for 90 minutes at 4°C, and centrifuged at 200 g for 1 minute to pellet the Sepharose slurry. The slurry was washed with 1 ml of IVT buffer, and pelleted by centrifugation as before. After two more wash and centrifugation cycles, 50 µl of 5x SDS gel loading dye (15% [v/v] β-mercaptoethanol, 15% [w/v] SDS, 50% [w/v] glycerol and 0.01% [w/v] bromophenol blue) was added to the pelleted Sepharose slurry and the suspension was placed in a boiling water bath for 2 minutes. The slurry was pelleted and the supernatant removed for analysis by PAGE.

2.11.5: Trichloroacetic acid precipitations.

The extent to which radioactive label had been incorporated into polypeptide products of *in vitro* translation reactions was assessed using trichloroacetic acid (TCA) precipitation. Two 1 µl aliquots of ^{35}S methionine labelled material generated from each translation reaction were placed on to individual Whatman 3MM filter paper discs. After the aliquots had dried, one of the discs was placed in a beaker containing 250 mls of ice-cold 10% (w/v) TCA to wash away unincorporated radioactivity. After

10 minutes, the TCA was discarded and the disks washed in fresh ice-cold 10% (w/v) TCA for a further 10 minutes. The filter disc was then removed and dried. This filter paper disc, which represented incorporated radioactivity, and the other filter paper disc, which represented total radioactivity, were placed in separate scintillation vials, covered with liquid scintillant and then placed in a scintillation counter where the radioactivity of the filter discs was measured. The efficiency by which the ^{35}S methionine had been incorporated into the polypeptide was determined by expressing the incorporated radioactive counts as a percentage of the total radioactive counts in each sample.

2.12: EXPRESSION OF POLYPEPTIDES IN BACTERIAL SYSTEMS.

2.12.1: Expression and purification of polypeptides using the Qiagen Express™ system.

Plasmids derived from pQE32 which contained DNA inserts to be expressed (pQE32.24 or pQE32.P2, see sections 9.2.1 and 8.3.2 respectively, for construction details) were transformed into *E. coli* M15 [pREP4] using the method described in section 2.5.2. The pQE32 plasmid carries a gene expressing ampicillin resistance, whereas plasmid pREP4 carries a gene expressing kanamycin resistance and so for maintenance of both plasmids, all plasmid propagations were performed in the presence of ampicillin (100 µg/ml) and kanamycin (25 µg/ml).

A 200 µl aliquot of an overnight culture of *E. coli* strain M15 [pREP4] transformed with a pQE series vector was inoculated into a 10 ml culture of 2YT medium in a 100 ml conical flask. The flask was placed in a 37°C shaking incubator until the OD₅₉₀ of the culture was 0.6 at which time expression of the desired protein

was initiated by the addition of IPTG to a final concentration of 10 mM. The induced culture was incubated for a further 3 hours after which the bacteria were pelleted by centrifugation at 2,500 g for 5 minutes. The bacterial pellet was resuspended using a vortex in 1.6 mls of Buffer B (8 M urea, 0.1 M sodium phosphate, 0.1 M Tris-HCl [pH 8.0]) and centrifuged at 5,000 g for 10 minutes. The supernatant was transferred to a new tube to which 400 µl of Ni-NTA binding slurry was added. The tube was vigorously agitated for 30 minutes at room temperature and then centrifuged at 200 g to pellet the slurry. The slurry was washed in buffer C (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris-HCl [pH 6.3]) by 3 rounds of pelleting, resuspension and centrifugation. In order to release the purified polypeptide from the slurry, 160 µl of buffer C containing 100 mM EDTA was added and the mix was incubated at room temperature for 2 minutes. The slurry was then pelleted and the supernatant containing the purified polypeptide was collected for later analysis by SDS-PAGE.

2.12.2: Expression of polypeptides using *E. coli* strain BL21.

pGEM1 derived transcription plasmids were transformed into *E. coli* BL21 in a method identical to that used for the transformation of *E. coli* strain TG2 (section 2.5.2). Cultures of the transformed *E. coli* BL21 were grown overnight in 2 mls of 2YT medium supplemented with ampicillin at a final concentration of 100 µg/ ml. A 200 µl aliquot of the culture was used to inoculate 10 mls of 2YT medium supplemented with ampicillin at 100 µg/ml contained in a 100 ml conical flask, and this was shaken at 37°C. When these cultures reached an OD₅₉₀ of 0.8, expression of the foreign plasmid was induced by the addition of IPTG to give a final concentration 10 mM. By inducing cultures with IPTG, the *E. coli* BL21 began to synthesise T7 RNA polymerase which then transcribed mRNA from the cloned insert. The mRNA was then translated using the bacterial protein synthesis machinery to produce the desired protein species. Cultures were incubated for a further 3 hours after which the bacteria

were pelleted by centrifugation at 2,500 g for 5 minutes. The bacterial proteins were solubilised by resuspending the pellet in 1 ml of isotonic lysis buffer (section 2.1.4). Insoluble material was pelleted by centrifugation at 10,000 g for 5 minutes and the resulting supernatant fraction was stored at -20°C until needed.

2.13: GENERATION OF ANTISERA IN LABORATORY ANIMALS.

2.13.1: Generation of antisera against the 24 kDa N protein related polypeptide.

Purified 24 kDa N protein species, expressed as described in section 2.12.1, was diluted in distilled water to an approximate concentration of 2 µg/ml (estimated by SDS-PAGE analysis) and then emulsified in an equal volume of Freund's complete adjuvant (Sigma Chemical Co.) by vigorous vortexing for 5 minutes. A 1 ml quantity of this emulsion was used to immunise a rabbit by sub-cutaneous injections at four sites. Three subsequent boosts of 24 kDa protein emulsified in Freund's incomplete adjuvant were performed at three week intervals. Two weeks after the final boost, blood was taken from the animal and the resulting serum was used in western blot analysis to determine whether the 24 kDa N species, in addition to the 43 kDa full-length N protein, was expressed in PVM-infected BSC1 cells.

2.13.2: Generation of antisera against the PVM P gene second ORF protein product.

An oligopeptide representing amino acids 45-62 of the predicted amino acid sequence of the PVM P gene second ORF protein product was synthesised and each linked to the carrier protein Keyhole limpet haemocyanin (KLH, Sigma Chemical Co.) according to the method described in Liu *et al.* (1979). KLH was dissolved in

phosphate buffered saline (PBS, 10 mM, pH 7.2) to a concentration of 15.6 mg/ml and dialysed against the same buffer for 24 hours. KLH (320 μ l) was activated by slow addition of 55 μ l of m-maleimidobenzoyl N-hydroxysuccinimide (MBS, Sigma Chemical Co.; 12 mg/ml, dissolved in dimethylformamide) and incubated for 30 minutes at room temperature. The activated KLH was applied to a 20 ml column of Sephadex G-25 (Pharmacia LKB) previously equilibrated with PBS (50 mM, pH 6) and 1 ml fractions were collected. The OD₂₈₀ of the fractions was measured and those that comprised the first peak, which contained activated KLH, were pooled (figure 5). The pooled fractions were added to an aqueous solution of each synthetic peptide (5 mg/ml). The reaction mixture was adjusted to pH 7.0 using 1 M HCl, and incubated for 3 hours at room temperature. This solution was emulsified with an equal volume of Freund's complete adjuvant (Sigma Chemical Co.) and used to immunise rats by subcutaneous injection. Following three boosting injections, separated by 3 week intervals, blood was taken from the animals and used in Western blot analysis to determine if the PVM P gene second ORF protein product was expressed in PVM-infected BSC1 cells.

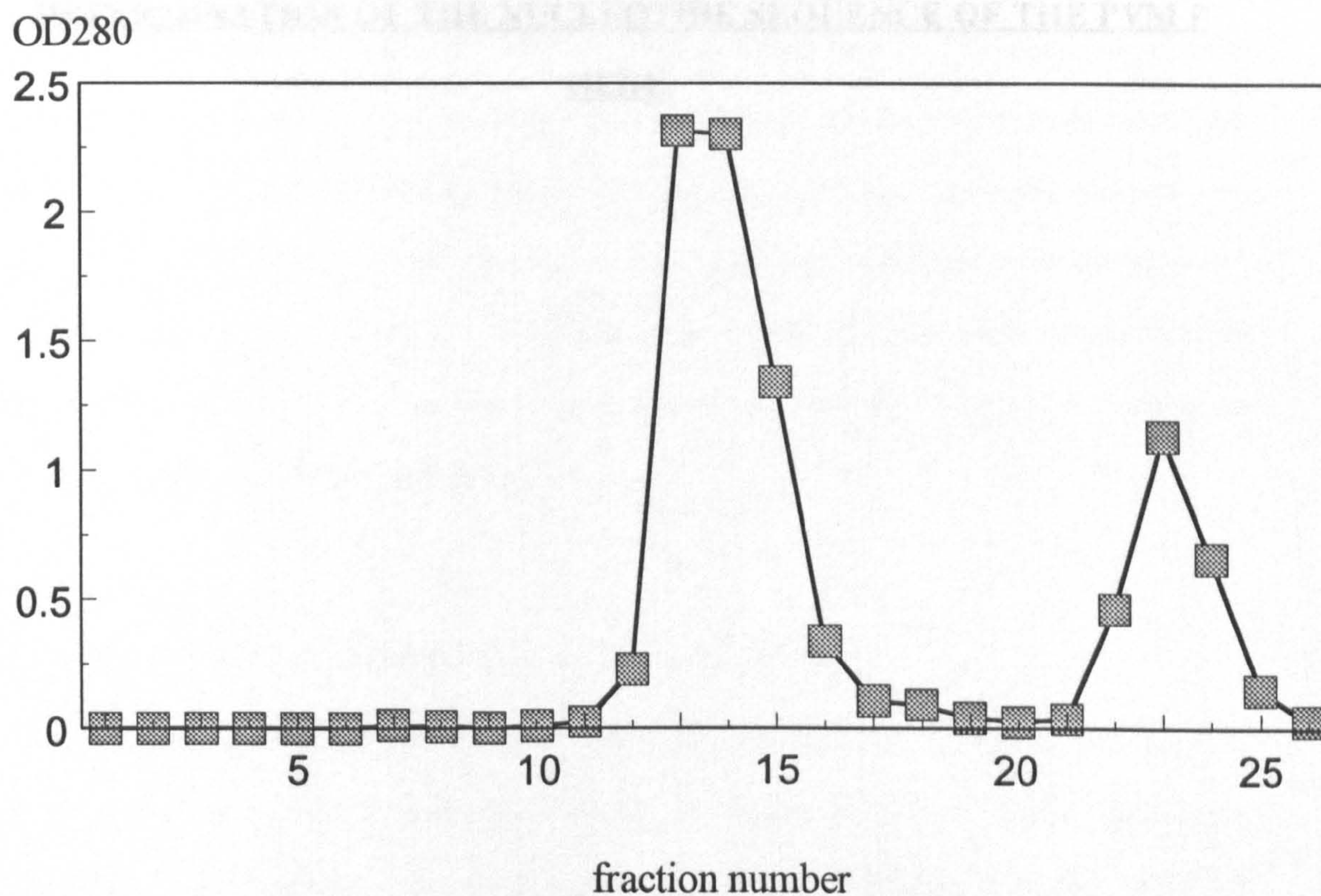


Figure 5. Graph to show the typical elution profile of MBS-activated keyhole limpet haemocyanin passing down a sephadex G-25 column. Fractions 13, 14 and 15, which contained MBS-activated KLH, were pooled for use in coupling to the synthetic peptide representing a region of the PVM P2 polypeptide. The second peak comprising fractions 22, 23 and 24 represents MBS.

CHAPTER 3

DETERMINATION OF THE NUCLEOTIDE SEQUENCE OF THE PVM P GENE.

3.1: CONSTRUCTION OF M13 VECTORS FOR SEQUENCING THE PVM P GENE.

Full-length P gene cDNA was excised from plasmid pUC13-P, and digested using a variety of restriction enzymes to generate a range of truncated cDNA fragments which are represented diagrammatically in figure 6. The restriction enzymes used to digest the P gene cDNA are shown on the figure. The resulting cDNA fragments were purified by agarose gel electrophoresis and fragment preparation and inserted into the multiple cloning site of the bacteriophage M13 derivatives mp18 and mp19 using standard molecular techniques. The resulting constructs were transformed into *E. coli* TG2. Single strand templates were prepared from these bacteriophage derivatives and used for nucleotide sequence analysis of the PVM P gene according to the procedures detailed in section 2.9.

3.2: THE NUCLEOTIDE SEQUENCE OF THE PVM P GENE.

The sequence of the PVM P gene mRNA and the predicted polypeptide sequence of the PVM P protein is shown in figure 8. The complete nucleotide sequence was determined by merging the sequence obtained from the range of bacteriophage M13 constructs containing P gene cDNA described in section 3.1. The PVM P gene is 903 nucleotides long and contains a long open reading frame (ORF) which starts from an AUG initiation codon at nucleotide positions 10 to 12 (AUG₁₀) and which finishes at a termination codon at positions 895 to 897. The size of the PVM P gene corresponds well with the estimated size of the mRNA observed in Northern blots (Chambers *et al.*, 1990a). The PVM P gene begins with the conserved PVM gene start consensus sequence (Chambers *et al.*, 1990a) and for this reason it is believed that the first nucleotides of the sequence presented here represents the exact 5' end of the P gene. The PVM P gene also ends with the PVM gene consensus stop sequence (Chambers *et al.*, 1990a) and is followed by a number of A residues.

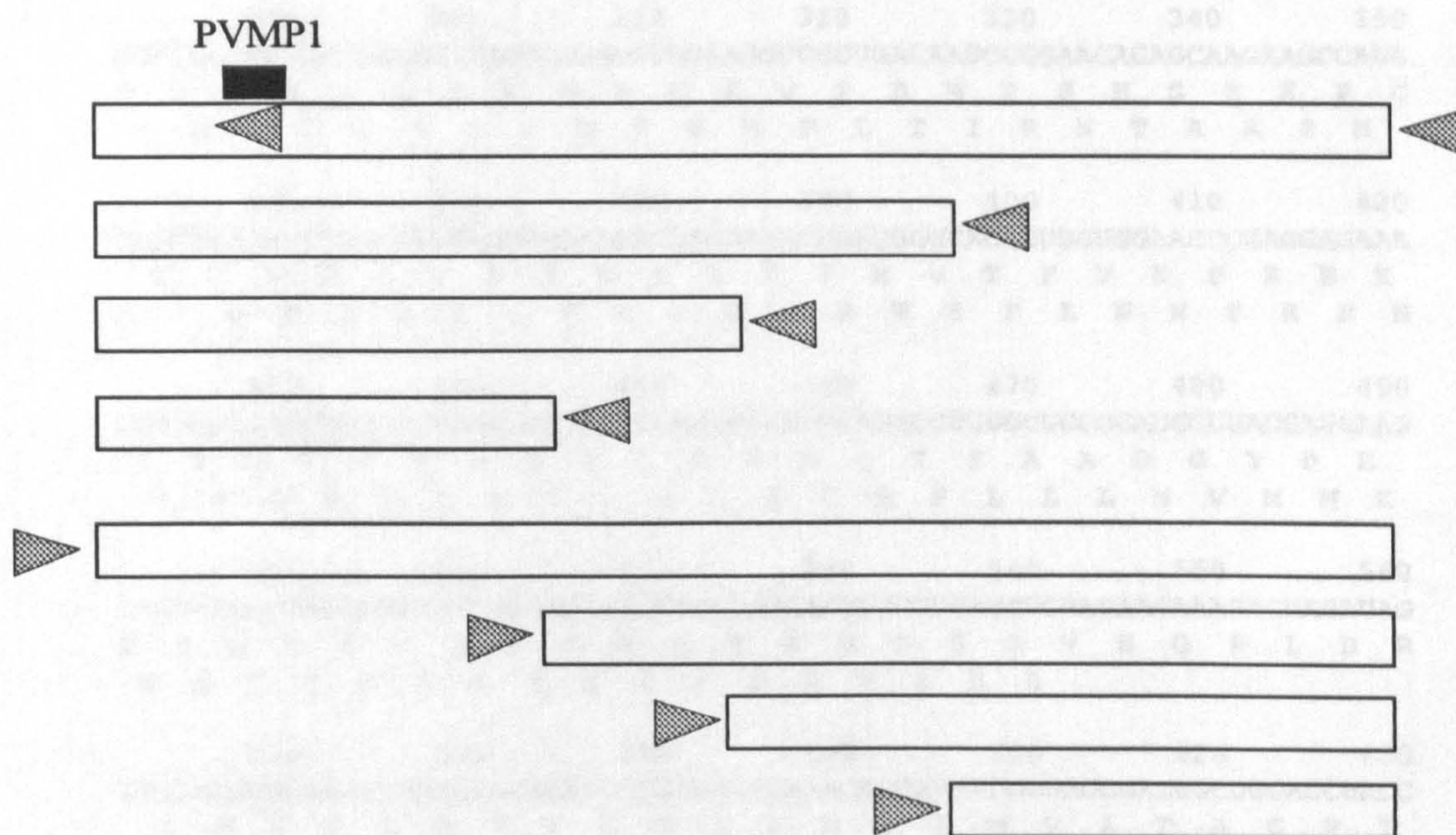
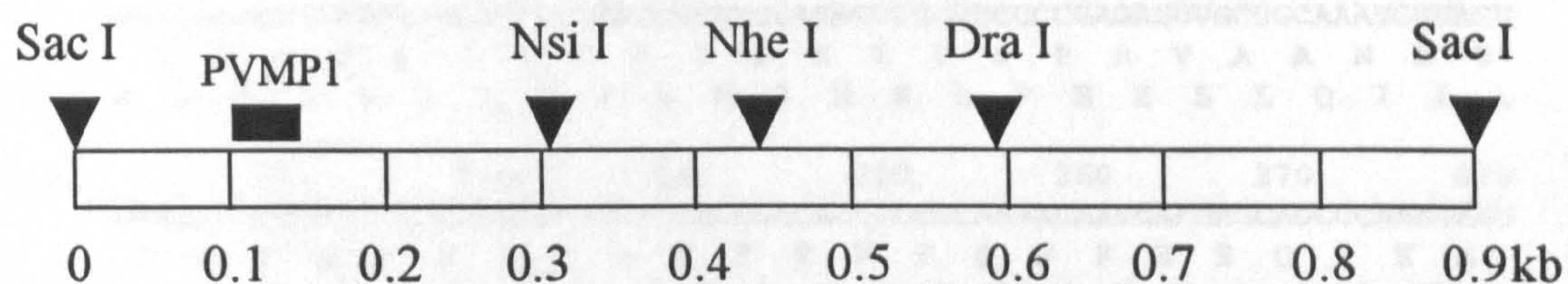


Figure 6. Region of the PVM P gene represented by M13 subclones constructed to determine the P gene nucleotide sequence. A fragment representing full-length P gene cDNA was removed from plasmid pUC13-P by digestion with restriction enzyme *Sac* I. This fragment was cleaved with restriction enzymes *Nsi* I, *Nhe* I and *Dra* I to create truncated cDNA segments which were ligated into the M13 derivatives mp18 and mp19 that had been cleaved with *Sma* I and *Sac* I. The nucleotide sequence of the segments was determined as described in section 2.9. In order to determine the 3' end of the PVM P gene, the -20 sequencing annealing primer (section 2.9) was substituted by the oligonucleotide primer PVMP1 (section 2.8). The position of the restriction enzyme sites used in these constructions are shown on the figure. The arrows within each cDNA segment denotes the direction in which the nucleotide sequence was read from each construct.

10	20	30	40	50	60	70
AGGAUAAUAUGGAGAAAUUCGCCCCGAAUUUGUUGGCGAGGAUGCUAACAAGAAGGCAGAGGAGUUUC						
M E K F A P E F V G E D A N K K A E E F						
80	90	100	110	120	130	140
UCAAAUAUGAUCCUUCUUUCGGAAAAACCACUAGCUGGUUAUACCGAACACUGCCACUCAUGUCACCAA						
L K H R S F P S E K P L A G I P N T A T H V T K						
M S P						
150	160	170	180	190	200	210
AUUAACAUGCCCCUAUAUUGCGUAGCUCAUCAAACUCCCUUCCCCGAGAGUUGCUGCAAUUCUUACU						
Y N M P P I L R S S F K L P S P R V A A N L T						
N I T C P L Y C V A H S N S L P R E L L Q I L L						
220	230	240	250	260	270	280
GAACCCUCUGCUCCCCUACCACUCCACCACCCACACCUCUCCCGAGAACAAGGAAGAGCAGCCCAAAGAGU						
E P S A P P T T P P P T P P Q N K E E Q P K E						
N P L L P L P L H H P H L P R T R K S S P K S						
290	300	310	320	330	340	350
CUGAUGUUGACAUUGAGACUAUGCAUGUCUGUAAGGUUCCUGACAAUCCGGAACACAGCAAGAAGCCAUG						
S D V D I E T M H V C K V P D N P E H S K K P C						
L M L T L R L C M S V R F L T I R N T A R S H						
360	370	380	390	400	410	420
CUGCUCAGAUGAUACCGAUACUAAGAAAACUAGGAAGCCGAUGGUCACCUUUGUGGAACCCGAGGAGAAA						
C S D D T D T K K T R K P M V T F V E P E E K						
A A Q M I P I L R K L G S R W S P L W N P R R N						
430	440	450	460	470	480	490
UUUGUCGGAUUGGGAGCUAGCUUGUACAGGGAGACCAUGCAGACCUUUGCUGCUGAUGGUUAUGAUGAAG						
F V G L G A S L Y R E T M Q T F A A D G Y D E						
L S D W E L A C T G R P C R P L L L M V M M K						
500	510	520	530	540	550	560
AAAGCAACCUAUCGUUUGAGGAGACUAACCAAGAGCCGGGUUCUUAUCUGUAGAACAAGACUAGAUAG						
E S N L S F E E T N Q E P G S S S V E Q R L D R						
K A T Y R L R R L T K S R V L H L						
570	580	590	600	610	620	630
AAUAGAGGAGAAUUGUCCUACAUAUAGGCCUUUUAACACCAUAAUGGUAGCGACUGCUGGACCUACC						
I E E K L S Y I I G L L N T I M V A T A G P T						
640	650	660	670	680	690	700
ACUGCUAGAGAUGAGAUUAGAGAUGCCCUUAUAGGCACUAGAGAAGAACUUAUUGAGAUGAUCAAGUCUG						
T A R D E I R D A L I G T R E E L I E M I K S						
710	720	730	740	750	760	770
ACAUCUUGACUGUCAUUGACAGAAUAGUGGCCAUGGAGAAGCUCAGAGAUGAGGAUUGCUCCAGAGCUGA						
D I L T V N D R I V A M E K L R D E E C S R A D						
780	790	800	810	820	830	840
CACUGAUGAUGGAUCAGCCUGUUAUUAACAGACAGAGCAAGGAUACUAGAUAAAGAUAGUGUCCAGCAAU						
T D D G S A C Y L T D R A R I L D K I V S S N						
850	860	870	880	890	900	
GCUGAAGAGGCUAAGGAAGAUUUGGAUGUUGAUGACAUCAUGGGCAUUAUUAUUAUAGUUAUUAU						
A E E A K E D L D V D D I M G I N F						

Figure 7. Nucleotide sequence of the PVM P gene shown in mRNA sense. The sequence of the predicted polypeptides encoded by the first ORF (initiation codon AUG₁₀) and the second ORF (initiation codon AUG₁₃₁) are shown in the single letter code. The numbers refer to the nucleotide positions in the gene.

Nucleotide sequence analysis of cDNA clones representing polycistronic mRNA species has been used to determine the sequence between the PVM P gene and the adjacent genes. This analysis indicates that neither of the regions bordering the PVM P gene contains sequences similar to the PVM consensus gene start and end sequences (Chambers *et al.*, 1990a, 1990b). For these reasons, it is believed that the nucleotide sequence presented here is complete. The 5' and 3' non-coding regions that are present at the beginning and end of the P gene are very short, being just 9 and 6 nucleotides in length respectively.

3.3: THE PREDICTED PRODUCTS OF THE PVM P GENE.

The PVM P gene long open reading frame (ORF) has the potential to encode a polypeptide 295 amino acids in length. Previous work by Ling and Pringle (1989a) identified the PVM P protein as having a molecular weight of 39 kDa by SDS-PAGE analysis. The assignment of this 39 kDa polypeptide as the PVM P gene product agreed with the results of hybrid arrested translation experiments in which a P gene cDNA clone prevented the expression of a 39 kDa polypeptide (Chambers *et al.*, 1990a). The expression of several other polypeptides was also arrested and the nature of these is discussed in section 4.3.2. This 39 kDa polypeptide product of the PVM P gene is referred to as P₃₉. The observed molecular weight determined for the PVM P₃₉ protein is not in close agreement with its predicted molecular weight which is 32,916 Da. Aberrant migration under SDS-PAGE seem to be a common feature of the P proteins of *Paramyxoviridae* family members, and possible reasons for this are discussed in section 4.3.1. The amino acid composition of the PVM P protein is shown in table 3.

non-polar amino acids	number in P protein	percentage of total residues
Ala	21	7.1
Val	15	5.1
Leu	19	6.4
Ile	17	5.8
Pro	24	8.1
Met	9	3.1
Phe	10	3.4
Trp	0	0
polar amino acids	number in P protein	percentage of total residues
Gly	11	3.7
Ser	21	7.1
Thr	23	7.8
Cys	5	1.7
Tyr	5	1.7
Asn	12	4.1
Gln	5	1.7
acidic amino acids	number in P protein	percentage of total residues
Asp	24	8.1
Glu	34	11.5
basic amino acids	number in P protein	percentage of total residues
Lys	21	7.1
Arg	15	5.1
His	4	1.4

Table 3. Amino acid composition of the PVM P protein as predicted from the PVM P gene nucleotide sequence.

From this table, the PVM P₃₉ protein can be seen to have a high overall percentage of acidic residues (19.6% of total polypeptide). The predicted pI of P₃₉ is 5.92. The acidic amino acids are not distributed evenly throughout the PVM P protein. In the amino-terminal 85 amino acids there are only 7 acidic residues whereas in the carboxyl-terminal 85 amino acids, 24 residues are acidic. Throughout the remaining central region, the acidic residues are evenly spaced. This uneven distribution of acidic residues will render the carboxyl-terminal negatively charged at neutral pH, and this polarity may be a functionally important feature of the PVM P₃₉ protein, especially since it is, by analogy with other *Paramyxoviridae* family members, believed to associate with many other virus components through electrostatic interaction. There are also large numbers of threonine and serine residues which, by virtue of their hydroxyl side groups, are potential phosphorylation sites. The serine and threonine residues appear to be evenly distributed throughout the length of the P protein.

The PVM P gene also contains a second smaller ORF which is 411 nucleotides in length. The coding region of the second ORF extends from an AUG codon located at nucleotide positions 131 to 133 of the PVM P gene (AUG₁₃₁), to a termination codon located at nucleotide positions 542 to 544. The second ORF has the potential to encode a polypeptide 137 amino acids in length with a predicted molecular weight of 15,932 Da. The amino acid composition of the predicted second ORF polypeptide product, designated P2, is shown in table 4.

From this table, P2 can be seen to have only three acidic residues which represents just 2.2% of the total polypeptide. This figure contrasts sharply with the acidic amino acid content of the PVM P protein which is 19.6% of the total polypeptide. This indicates that P2 does not have any regions that at physiological pH would be negatively charged, suggesting that this polypeptide may not associate electrostatically with a positively charged domain of another polypeptide moiety. The P2 polypeptide is composed of a high proportion of basic amino acids with an overall

non-polar amino acids	number in P2 protein	percentage of total residues
Ala	6	4.4
Val	4	2.9
Leu	28	20.4
Ile	5	3.6
Pro	14	10.2
Met	7	5.1
Phe	1	0.7
Trp	3	2.2
polar amino acids	number in P2 protein	percentage of total residues
Gly	2	1.5
Ser	12	8.8
Thr	8	5.8
Cys	5	3.6
Try	2	1.5
Asn	6	4.4
Gln	2	1.5
acidic amino acids	number in P2 protein	percentage of total residues
Asp	1	0.7
Glu	2	1.5
basic amino acids	number in P2 protein	percentage of total residues
Lys	6	4.4
Arg	17	12.4
His	6	4.4

Table 4. Amino acid composition of the PVM P2 polypeptide as predicted from the PVM P gene nucleotide sequence.

content of 21.2%. The basic amino acids are not spread evenly throughout the length of the second ORF polypeptide. They are predominantly absent from the amino-terminus and there is a small clustering at the immediate carboxyl-terminal end of the polypeptide. Despite the low number of acidic amino acids and high numbers of basic residues, the predicted pI (6.7) of this polypeptide is close to neutral.

3.4: DISCUSSION

3.4.1: The PVM P gene.

The PVM P gene is 903 nucleotides in length and contains two open reading frames (ORFs). The first ORF, which is 885 nucleotides long, begins at an AUG codon located at nucleotide positions 10 to 12 (AUG₁₀) and runs until the UAG stop codon located at nucleotide positions 895 to 897. This ORF has the potential to encode a polypeptide 295 amino acids in length. The second ORF is 411 nucleotides in length, beginning on the third AUG codon of the P gene located at nucleotide positions 131 to 133 (AUG₁₃₁) and ending at the UAG codon at nucleotide positions 542 to 544. This ORF has the potential to encode a polypeptide 137 amino acids in length.

3.4.2: Comparison of the PVM P protein with the P proteins of other *Paramyxoviridae* family members.

Significant homology between the amino acid sequence of the PVM P₃₉ protein and the P protein of any other negative sense, single stranded, RNA virus, excepting HRSV, BRSV and TRTV could not be detected. This finding is in agreement with many previous studies which have failed to find amino acid homology between the

HRSV P protein and the P proteins of several *Paramyxoviridae* family members (Luk *et al.*, 1986). Spriggs and Collins (1986) identified a low level of homology between the HRSV P protein and a central domain of the PIV-3 and measles virus P proteins, but it is not known if this homology is significant.

By analogy with the other *Paramyxoviridae* family members, the nucleocapsid complex of PVM, which is composed of the N, P, and L proteins, is believed to harbour an RNA dependent RNA polymerase which directs the synthesis of viral RNAs during transcription and replication. Recently, in support of this analogy, studies have shown that HRSV possesses similar transcriptional machinery to other non-segmented, negative stranded RNA viruses (Huang *et al.*, 1993). In accordance with the overall functional similarity of the PVM nucleocapsid complex to these other nucleocapsid complexes, the PVM P protein would be expected to show at least some functional similarities to these other P proteins. The P proteins of many *Paramyxoviridae* family members show homology to one another in both the amino and carboxyl-terminal regions but show little homology in the central region (Matsuoka *et al.*, 1991). Examination of hydropathy plots of the P proteins of several members of the family *Paramyxoviridae* has not revealed the presence of obvious hydrophobic or hydrophilic motifs common to these P proteins. However the P (or NS) proteins of all *paramyxovirus* and *morbillivirus* genera members are considerably larger than the pneumovirus P protein and this feature complicates the identification of amino acid and structural homologies that may exist between these proteins.

3.4.3: Comparison of the PVM P protein with other pneumovirus P proteins.

A significant difference between the P proteins encoded by the PVM, HRSV and BRSV phosphoprotein genes is their length. The PVM P₃₉ protein is 295 residues long whereas the HRSV and BRSV P proteins are both 241 residues in length. The amino acid homology between the P proteins of these viruses is not high, with PVM P₃₉ and HRSV P proteins (HRSV A2 strain) having 35.3% overall identity and PVM P₃₉ and BRSV P proteins having 35.6% overall identity. The amino acid sequence of the PVM P protein and the P proteins from three strains of HRSV are compared in figure 8. The homology that is seen between the PVM P₃₉ protein and the HRSV and BRSV P proteins is not evenly distributed throughout the length of the proteins, but rather is concentrated into two regions of good homology separated by a region of poor homology. This homology distribution can be seen in figure 9. During sequence analysis of the pneumovirus P genes, all the gap substitutions introduced to generate alignment were inserted into the poorly homologous region, a feature that highlights the lack of similarities in this internal region of the pneumovirus P proteins. The first region of good homology is located at the extreme amino-terminus of these proteins and is only 17 amino acids in length. In this short region there are 14 conserved residues between the PVM P₃₉ protein and both the HRSV (strain A2) and the BRSV P proteins. The second region of high homology between these P proteins is located between residues 181 and 249 on the PVM P₃₉ protein sequence and 134 and 201 on the HRSV and BRSV P protein sequences. The percentage identities for this region are 64% for PVM and HRSV and 67% for PVM and BRSV. These figures increase to 84% and 86% respectively when similarities are also considered. Comparison between the amino acid sequence of the PVM and TRTV P proteins reveals that in TRTV, this second region of homology is also conserved, although the first region of homology at the amino-terminus is not (Ling *et al.*, in preparation). The percentage identity seen between the P

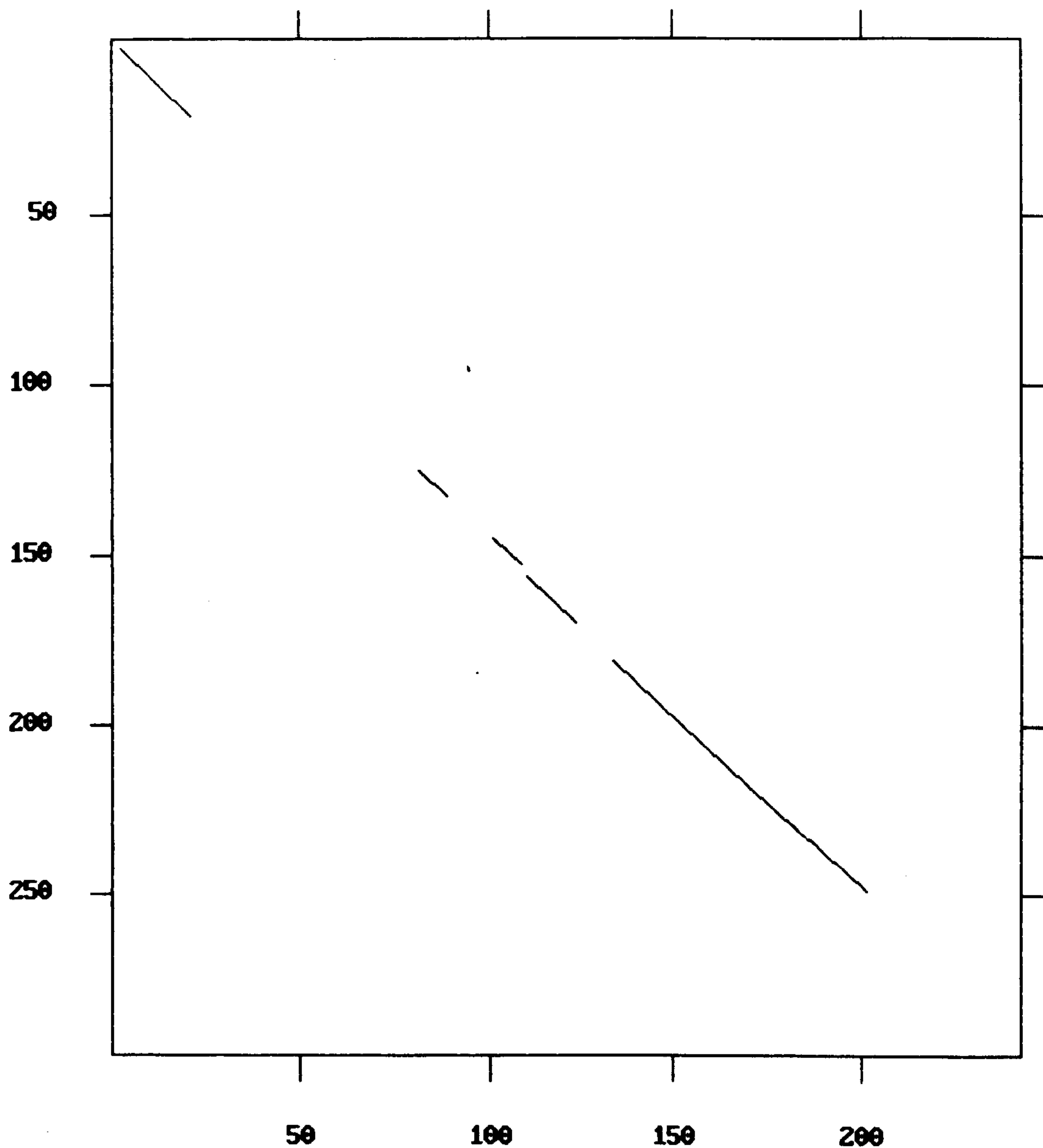


Figure 9. Matrix plot of identities between the PVM P protein (vertical axis) and the HRSV P protein (horizontal axis). The plot was obtained using Microgenie software (Beckman Instruments Inc.). The plot indicates that the identity that exists between these two proteins is not distributed evenly throughout their length, but is concentrated into two regions of good identity separated by a region of poor identity. The numbers along the axes refer to the amino acids of each protein.

proteins of PVM and TRTV in this region is 69%, rising to 82% upon consideration of similarities. The high level of homology that is seen between all four pneumoviruses in this region suggests that a functional domain may be contained here. The level of amino acid identity that is seen in this region is probably responsible for the observation that the PVM P₃₉ protein is serologically related to the HRSV and BRSV P protein (Gimenez *et al.*, 1984; Ling and Pringle, 1989a). In the poorly homologous amino-terminal region of the PVM P₃₀ protein, which spans residues 18 to 180, there are only 38 identities with the HRSV P protein, and this corresponds to a percentage of just 23%. Interestingly, in this region there is a high degree of conservation between serine and threonine residues, the potential phosphorylation sites, with 13 of the 38 identities in this region being either serine or threonine conservations. The percentage homology between the PVM P₃₉ protein and the BRSV P protein in this poorly homologous region is also extremely low, at 20.6%. The same distribution of homology that is found between the PVM and HRSV P proteins is also found between the two HRSV subgroups. Johnson and Collins (1990) identified two relatively large conserved domains located between residues 1-58 and 86-241 that between the two subgroups displayed an overall identity of 96%. However, the intervening region which separated these conserved domains was divergent showing an identity of only 52% between subgroup A and B isolates. Mallipedi and Samal (1992) demonstrated that the corresponding region of the BRSV P protein showed poor identity (37%) with the HRSV P protein of either subgroup over this region. This region lies within the portion of the HRSV and BRSV P proteins that displays poor homology with the corresponding region of the PVM P₃₉ protein. It is possible that this poorly conserved region that is present on all these three pneumovirus P proteins does not contain a functional domain which requires strict conservation of amino acid residues, and is consequently tolerant to non-conservative amino acid changes.

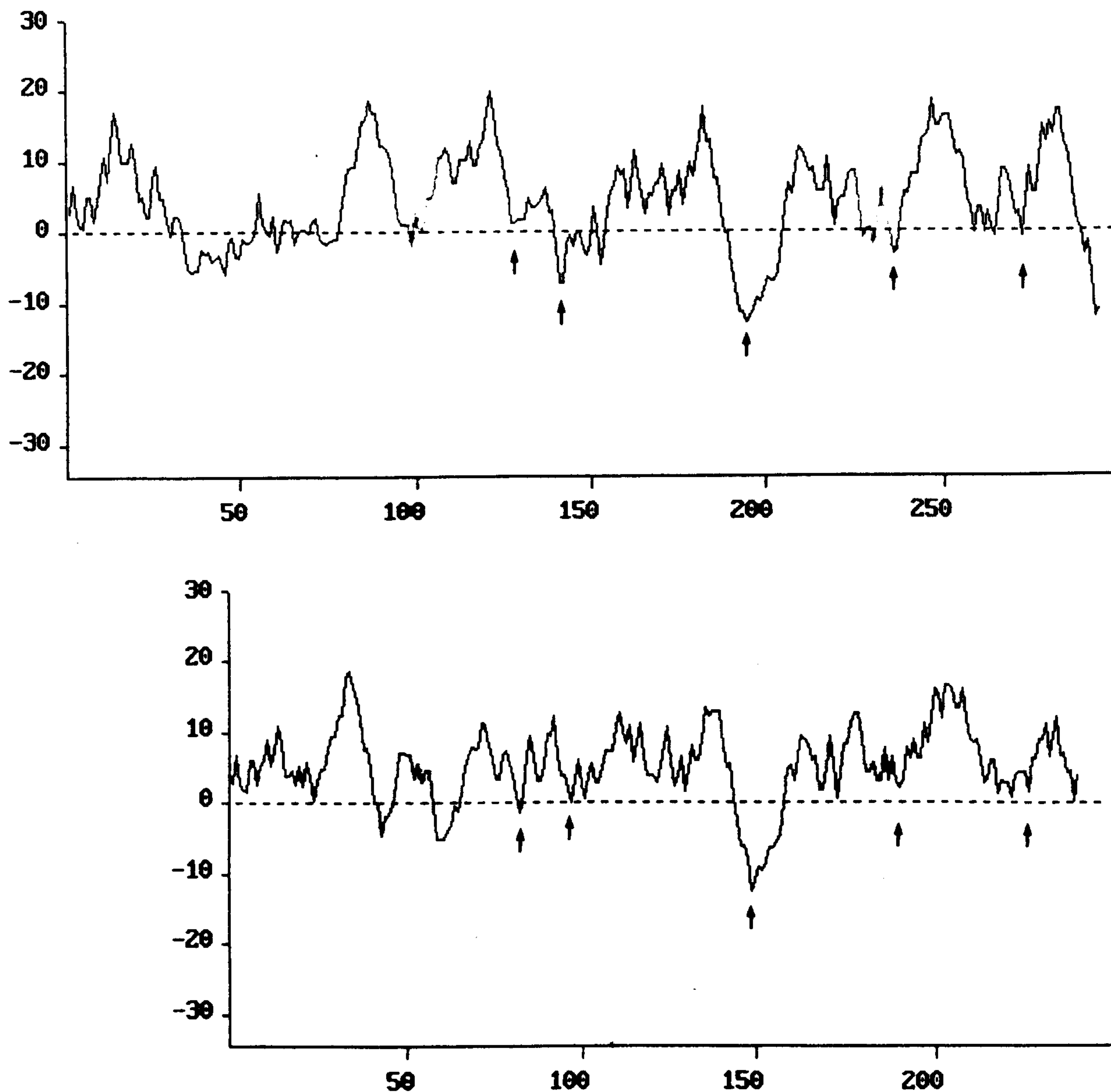


Figure 10. Hydropathy plots of PVM P protein (upper plot) and (B) HRSV P protein (lower plot). The plots were generated using the procedure of Hopp and Woods (1981) using a window of 10 amino acids. The difference in the sizes of these P proteins mean they do not align over their entire lengths. However, by removing a 50 amino acid long segment from the PVM P protein hydropathy plot between residues 30 and 80, the two proteins show a good hydropathic similarity throughout their lengths. Arrows indicate some regions of hydropathic alignment.

In addition to performing a comparison between the amino acid sequence of the pneumovirus P proteins, their corresponding predicted hydropathy plots were also compared. A comparison between plots of the HRSV P and the PVM P₃₉ proteins, shown in figure 10, is at first glance misleading because the difference in their sizes means that they do not align over their entire lengths. However by removing a 50 amino acid segment from the PVM P₃₉ protein hydropathy plot between residues 30 and 80, the HRSV and PVM P₃₉ protein hydropathy plots show a high degree of similarity. The resulting hydropathy alignment shows a matching of all hydrophilic peaks and all hydrophobic troughs. This degree of similarity shows that despite the low overall amino acid identities that these two proteins share, they show conservation of hydropathy, and presumably structural similarities, throughout their entire lengths. The portion of the PVM P₃₉ protein removed to create the alignment represented a small hydrophobic region which is not present on the HRSV hydropathy plot. This portion of PVM sequence lies completely in the region of the PVM P₃₉ protein that shows no significant homology with the other pneumovirus P proteins and this supports the proposal that the size difference between the PVM and RSV P proteins is due to either loss or gain of sequence in this region.

3.4.4: The PVM P gene second ORF.

The position of the PVM P gene second ORF coincides with the region of the PVM P protein that is poorly homologous to the BRSV and HRSV P proteins, with the carboxyl-terminal end of the second ORF product lying almost exactly at the end of the region of poor homology. The termination codon for the PVM second ORF product is 8 nucleotides upstream of the codon that encodes the first residue of the large region of high conservation. The region of poor homology includes the entire length of the sequence corresponding to the second ORF. The apparent lack of need

for strict conservation of amino acid residues in this region has permitted changes at the nucleotide level to occur and alter the sequence of both the first and the second open reading frames. Included in these alterations has been the removal of all the stop codons in the second ORF for the length of the variable region, and the subsequent creation of a substantial coding region in the second ORF. There are 17 stop codons in the corresponding region of the third reading frame. The probable need for more strict amino acid conservation in the immediately adjacent homologous region of the PVM P₃₉ protein has presumably prevented extension of the second ORF coding region past the boundary of the variable and homologous regions, thereby limiting the size of the second ORF. It would be interesting and informative to sequence the P genes of other PVM strains to see if the continuity of the second ORF is maintained. If P genes of other PVM strains were found to show differences in this region, but still possess the intact second ORF, then this would support the suggestion that the second ORF polypeptide performs a function that is advantageous to the replication of the virus. The P genes of BRSV, HRSV and TRTV do not contain an equivalent second extensive ORF.

3.4.5: Comparison of the PVM P protein sequence with a protein sequence database.

A more wide ranging search on a protein sequence database failed to identify any protein sequence which displayed amino acid sequence homology over a sufficiently large region to suggest definite relatedness. This search, however, did reveal a low level of amino acid homology between several short stretches of the PVM P₃₉ protein and correspondingly short stretches of proteins that were related to one another by their function as protein kinases. This homology search using the SWEEP program searching the GENBANK database used small 15 amino acid sections of the

PVM P₃₉ protein and compared them to small sections of database protein sequence, thus it could identify small regions of homology that would be often be lost if larger regions of sequence were being compared. The database protein sequence that each PVM P₃₉ protein section was homologous to was not always the same protein kinase. The short stretches of homology that were identified were spread virtually throughout the entire length of the P₃₉ protein amino acid sequence, with only amino acid sections 106-120, 210-225 and 240-295 showing no homology to a protein kinase amino acid sequence. There has been no report as yet of the P protein of *Paramyxoviridae* family members having any auto-phosphorylation activity and so the relevance of this observed homology is unclear. It is possible that the similarities reflect motifs found in phosphorylated proteins. However due to the lack of amino acid homology between PVM P protein and the P proteins of other *Paramyxoviridae* family members, it is possible that the PVM P protein exhibits fundamental functional differences to these other P proteins.

3.4.6: Comparison of the PVM P2 protein with the P gene products of other *Paramyxoviridae* family members.

A database search revealed no homology between the PVM P gene second ORF product, P2, and any protein product of another *Paramyxoviridae* family member. The predicted P2 polypeptide does not possess a cysteine rich region that is a common feature of the V protein which is encoded by the P genes of most *Paramyxoviridae* family members.

A potential mRNA editing site that could be utilised by the PVM RNA dependent RNA polymerase to switch between the first and second ORFs during P gene transcription was not identified. Mallipedi and Samal (1992) identified a region on BRSV P gene mRNA that resembled the consensus insertion sequence of

Paramyxoviridae family members where the viral polymerase stutters and insertion occurs (Vidal *et al.*, 1990b). A number of P gene cDNA clones were sequenced and in one instance, an insertion of three nucleotides was observed at the position of the proposed insertion site. No such insertions have been identified in any of the PVM P gene cDNA clones that have been sequenced, and a site that resembles the insertion consensus sequence is not present on the PVM P gene cDNA. However the possibility that the PVM P gene exhibits some form of mRNA editing should not be ruled out as it is possible that the PVM P gene uses an editing mechanism that is different to the other *Paramyxoviridae* family members.

CHAPTER 4

CHARACTERISATION OF THE *IN VITRO* EXPRESSION PRODUCTS OF THE PVM P GENE.

4.1: CONSTRUCTION OF T7 TRANSCRIPTION PLASMID pGEMP.

A cDNA fragment representing the full-length PVM P gene was removed from plasmid pUC13-P by restriction enzyme digestion, and inserted into the *Sac* I site of the T7 transcription plasmid, pGEM1 (Promega Corporation). However, as a consequence of the procedure used to obtain a full-length cDNA clone of the PVM P gene (section 2.8), a long region of cytosine residues were located at the 5' end of the P gene cDNA sequence. Attempts to translate mRNA transcribed from this pGEM1 derived plasmid were unsuccessful and this was believed to be due to the presence of this poly-C leader region. No convenient restriction enzyme sites were present to enable removal of this poly-C region by simple restriction enzyme digestion. In order to construct a full-length P gene cDNA clone without this leader sequence, PCR amplification on P gene cDNA was used to generate a fragment representing P gene nucleotides 1 to 139, thus removing the poly-C leader region. The oligonucleotide primers used for the amplification of the P gene cDNA fragment were designated PVMP2 (hybridises to P gene nucleotides 1-23; CGAGCTCGAGGATAAATATGGAGAAATTCGC) and PVMP1 (hybridises to P gene nucleotides 115-139; GGTGACATGAGTAGTGGCAGTGTTCGG). The following construction details of plasmid pGEMP are shown in figure 11. The successfully PCR amplified fragment was blunted using Klenow fragment and then cleaved with restriction enzyme *Acc* I. The cDNA segment was ligated into *Acc* I and *Sac* I cleaved pUC13-P. The full-length P gene cDNA fragment was cleaved from this plasmid by digestion with restriction enzyme *Sac* I and then inserted into the *Sac* I site of pGEM1. Correct orientation of the fragment was confirmed by restriction enzyme digestion of the ligated plasmid. This transcription plasmid was designated pGEMP.

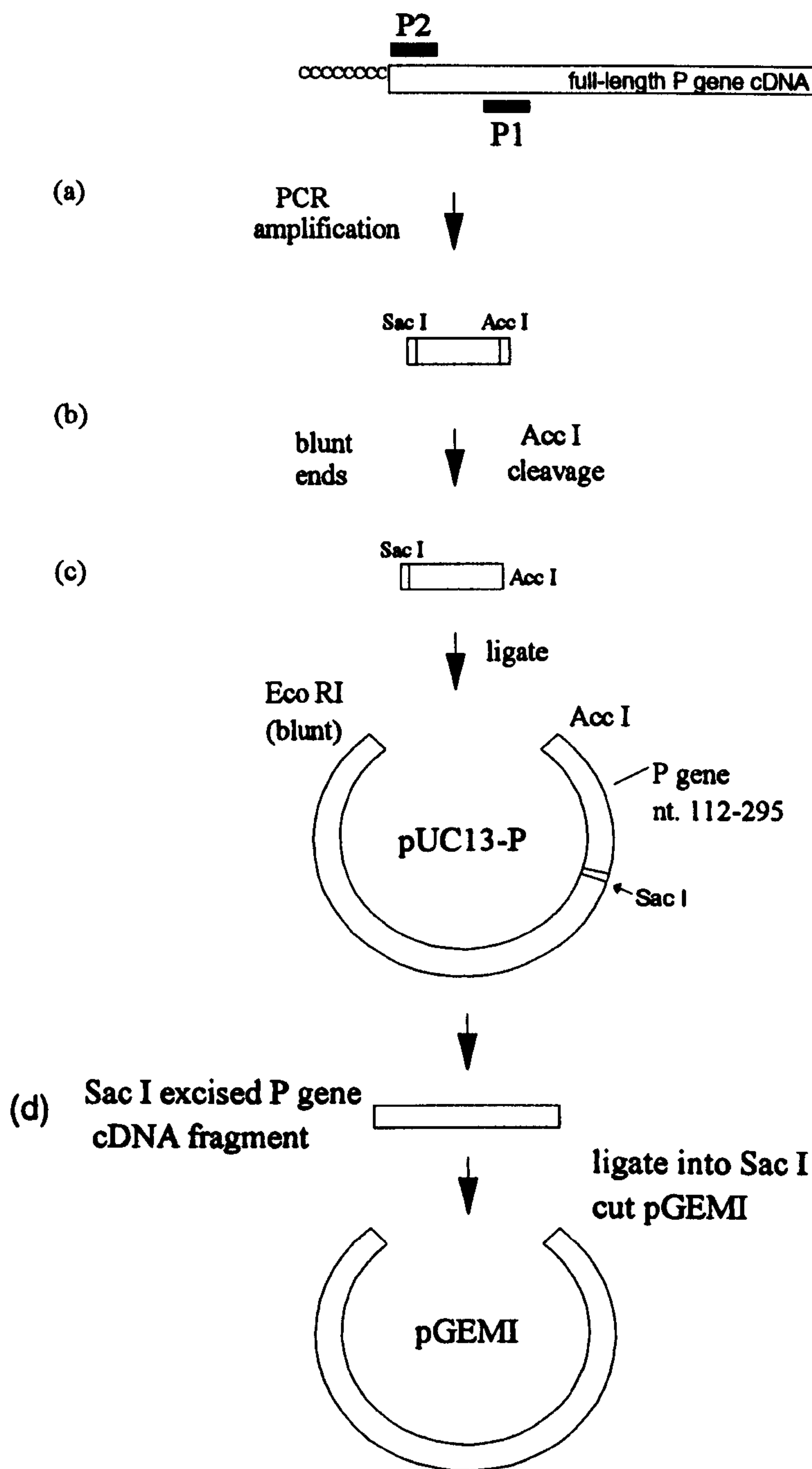


Figure 11: Construction of transcription plasmid pGEMP. A cDNA fragment representing the full-length PVM P gene was removed from pUC13-P (plasmid construction details in section 4.1). (a). A cDNA segment representing nucleotides 1 to 139 of the PVM P gene was amplified by PCR using oligonucleotide primers PVMP1 and PVMP2 thus removing the 5' poly(dC) tail from the full-length P gene cDNA clone. (b). The fragment was blunted using Klenow and digested using restriction enzyme *Acc I* at a site located within primer PVMP1. (c). The cleaved fragment was ligated into *Eco RI* (Klenow blunted) and *Acc I* cleaved pUC13-P thus reforming a full-length cDNA copy of the PVM P gene that did not possess a 5' poly(c) tail. (d). The full-length P gene cDNA was removed from pUC13-P using restriction enzyme *Sac I*, and this fragment was ligated into *Sac I* cut pGEMI thus deriving pGEMP.

4.2: *IN VITRO* EXPRESSION OF PVM P PROTEIN AND P PROTEIN RELATED SPECIES.

4.2.1: *In vitro* expression of the PVM P protein.

RNA transcribed from pGEMP was translated in a rabbit reticulocyte lysate system and the ^{35}S methionine labelled polypeptide products were analysed by SDS-PAGE. An autoradiograph of a gel on which these radiolabelled polypeptide products were electrophoresed is shown in figure 12. The primary polypeptide product of the PVM P gene cDNA is a 39 kDa species which represents the PVM P₃₉ protein. In addition, at least 3 additional polypeptide species are synthesised, having molecular weights of approximately 37 kDa, 26 kDa and 23 kDa. A fourth polypeptide with a molecular weight of 16 kDa is also seen following long exposure of the autoradiograph. These species, referred to as P₃₇, P₂₆, P₂₃, and P₁₆ respectively, are expressed at a far lower abundance than the 39 kDa species, and this can be seen in figure 12. The nature of these species is discussed in section 4.3. When subjected to SDS-PAGE, the 39 kDa species expressed *in vitro* from pGEMP transcribed mRNA comigrates exactly with the PVM-infected BSC1 cell polypeptide species previously identified as the PVM P protein (Cash, 1979, Ling and Pringle, 1989a).

4.2.2: Construction of pGEMP derived transcription plasmids.

As described in section 4.2.1, the mRNA transcribed from full-length P gene cDNA directs the synthesis of four smaller polypeptide species with molecular weights of 37 kDa, 26 kDa, 23 kDa and 16 kDa. In order to determine how these species arise, a series of T7 transcription plasmids containing 5' truncated PVM P gene cDNA inserts were constructed.

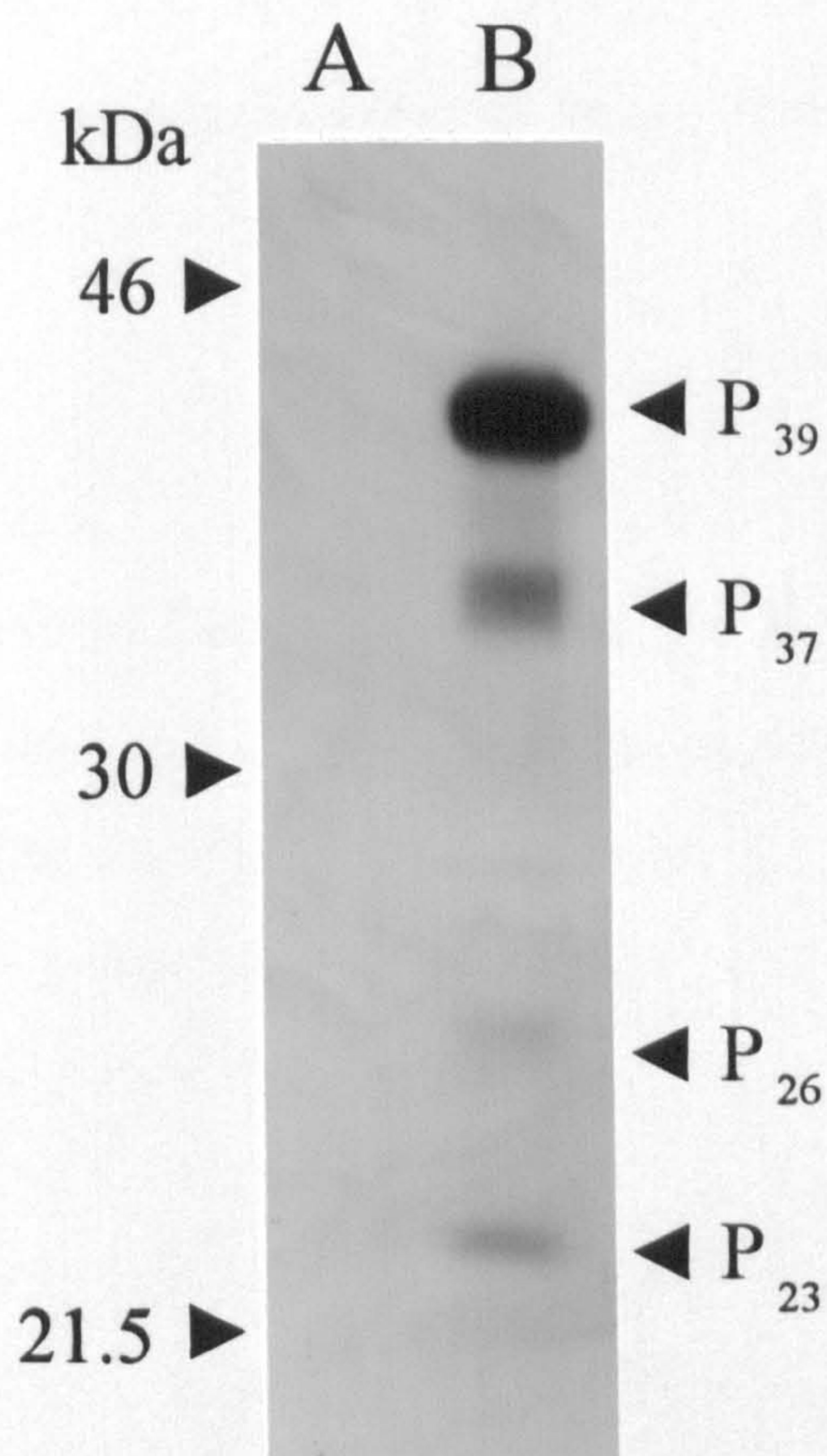


Figure 12. The full-length P gene cDNA directs the *in vitro* expression of the full-length P protein (P₃₉) and three smaller polypeptides with molecular weights of 37 kDa (P₃₇), 26 kDa (P₂₆) and 23 kDa (P₂₃). mRNA transcribed from the full-length P gene cDNA insert within pGEMP was used to programme a reticulocyte lysate system and the radioactively labelled expression products were subjected to SDS-PAGE and auto-radiography. Lane A: No mRNA. Lane B: pGEMP transcribed mRNA. The positions of molecular weight markers are indicated on the left, and the polypeptides expressed from pGEMP transcribed mRNA are marked on the right.

The full-length P gene cDNA fragment excised from pGEMP was truncated by restriction enzyme digestion and inserted into pGEM1 to generate the vectors pGEMP37, pGEMP26 and pGEMP23. The construction details of these plasmids are shown diagrammatically in figure 13. The first ATG codons downstream of the T7 promoter on each of these four pGEM1 derived transcription vectors (ATG₁₀, ATG₁₄₈, ATG₃₀₁, and ATG₃₉₁) correspond to the first, second, third and fourth AUG codons respectively of the PVM P gene long ORF.

4.2.3: *In vitro* expression of P protein related species.

Radioactively labelled P protein related polypeptides were translated *in vitro* using rabbit reticulocyte lysate programmed with mRNA transcribed from plasmids pGEMP, pGEMP37, pGEMP26 and pGEMP23. The radiolabelled translation products were analysed by SDS-PAGE and an autoradiograph of the resulting gel is shown in figure 14.

In pGEMP the first initiation codon that is in-frame with the P gene long open reading frame is ATG₁₀. In the construction of pGEMP37, ATG₁₀ was removed and the first in-frame ATG on this plasmid became ATG₁₄₈ which corresponds to the second AUG codon on the P gene mRNA that is in-frame with the P gene long ORF. The plasmid pGEMP37 directs the synthesis of the P₃₇, P₂₆, P₂₃ and P₁₆ polypeptides, which comigrate with the species expressed from pGEMP, but pGEMP37 does not express the full-length P protein (figure 14). This suggests that AUG₁₀ on the mRNA transcribed from pGEMP is responsible for initiation of the full-length P protein, since removal of the corresponding ATG in the case of pGEMP37, prevents synthesis of the full-length P protein. In the construction of plasmid pGEMP26, ATG₁₄₈ was removed and mRNA transcribed from this plasmid directs the synthesis of the P₂₆, P₂₃

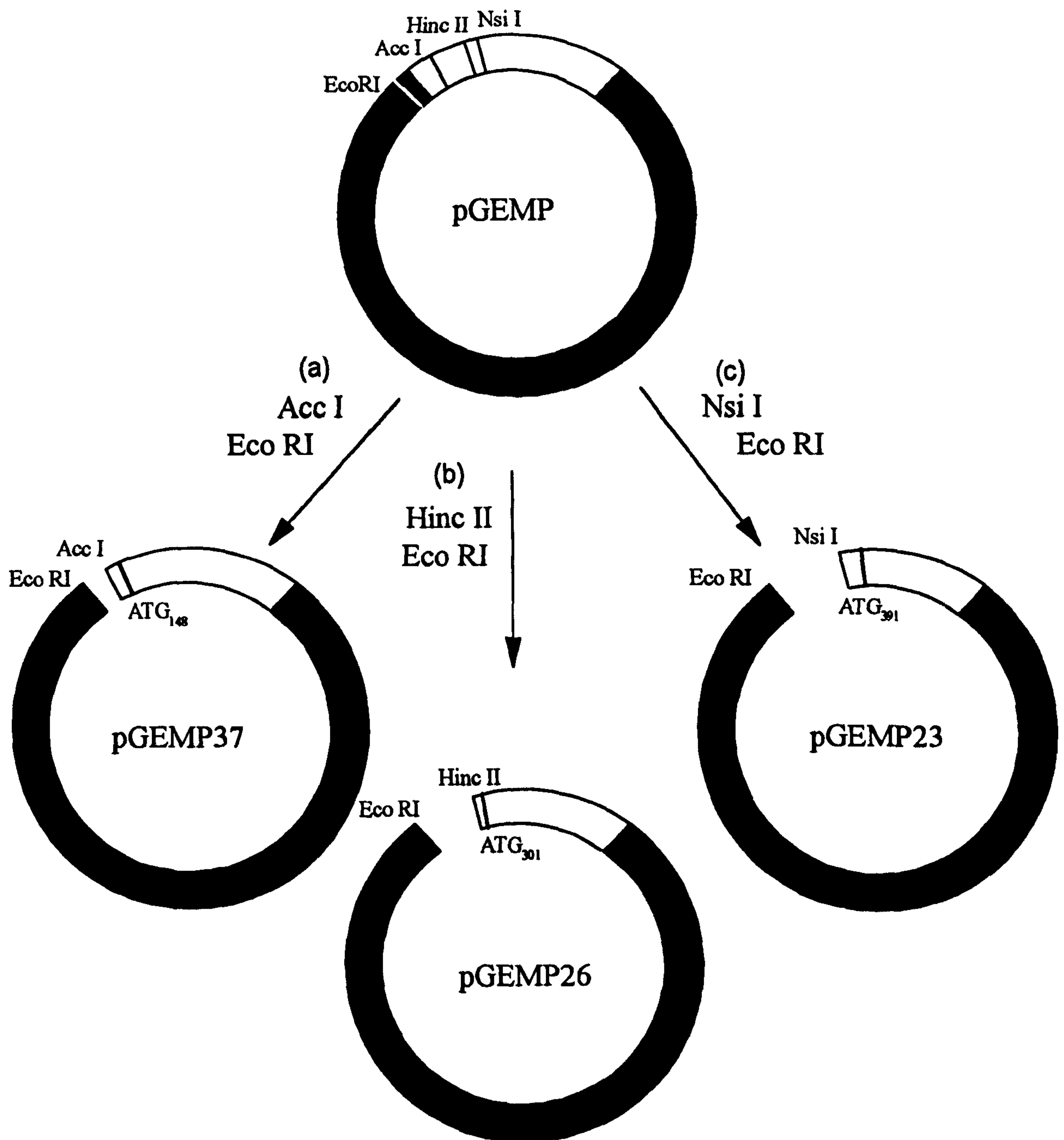


Figure 13. Construction of T7 transcription plasmids pGEMP37, pGEMP26 and pGEMP23.

(a). Construction of pGEMP37: pGEMP was cleaved with restriction enzymes *Acc* I and *Eco* RI to remove nucleotides 1-112 of the PVM P gene insert. The *Eco* RI site is present in the pGEMP multiple cloning site. The ends were blunted using Klenow and religated to create pGEMP37.

(b). Construction of pGEMP26: pGEMP was cleaved with restriction enzymes *Hinc* II and *Eco* RI to remove nucleotides 1-289 of the PVM P gene insert. The ends were blunted using Klenow and religated to form pGEMP26.

(c) Construction of pGEMP23: pGEMP was cleaved with *Nsi* I and *Eco* RI enzymes to remove nucleotides 1-303 of the P gene insert. The ends were blunted using Klenow and religated to give the plasmid pGEMP23. The first initiation codon of the P protein ORF is marked on each plasmid.

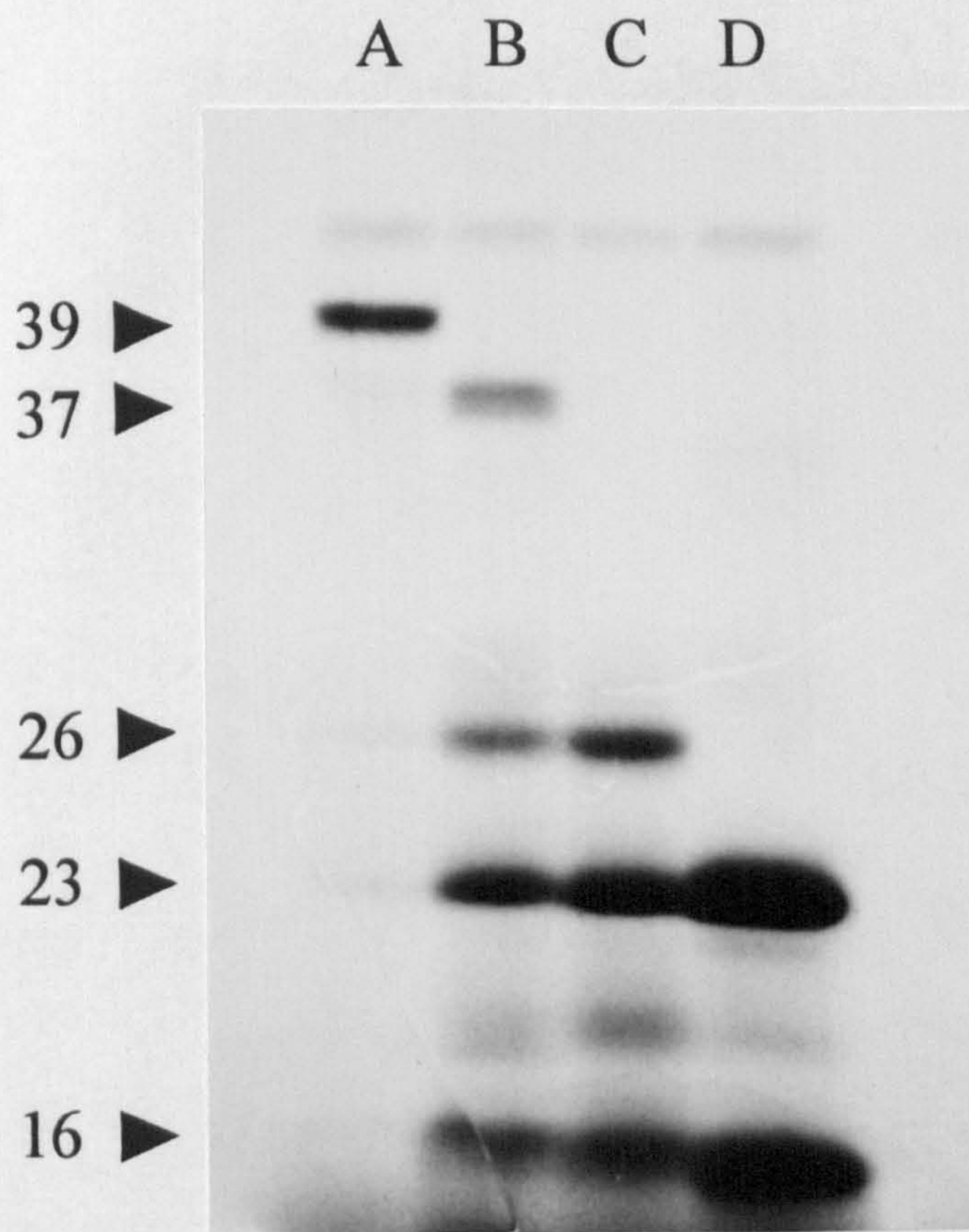


Figure 14. Identification of AUG codons used to initiate translation of five *in vitro* synthesised P protein related polypeptides. Radioactively labelled P protein related polypeptides were translated *in vitro* using rabbit reticulocyte lysate programmed with mRNA transcribed from plasmids pGEMP, pGEMP37, pGEMP26 and pGEMP23. Radioactively labelled translation products expressed from mRNAs transcribed from these plasmids were subjected to SDS-PAGE. The gel was loaded in the following order. Lane A: pGEMP directed products. Lane B: pGEMP37 directed products. Lane C: pGEMP26 directed products. Lane D: pGEMP23 directed products. Following electrophoresis the gel was fixed, dried and autoradiographed in order to visualise the labelled polypeptides.

and P₁₆ polypeptides but not the P₃₇ species. Thus the corresponding initiation codon, AUG₁₄₈, on the P gene mRNA transcribed from pGEMP37, and also pGEMP, is responsible for initiation of the P₃₇ protein.

Transcription plasmid pGEMP23 has lost ATG₃₀₁ and directs the synthesis of the P₂₃ and P₁₆ species. Loss of ATG₃₀₁ has prevented expression of the P₂₆ species and so the corresponding codon AUG₃₀₁ on the mRNA transcribed from pGEMP26 appears to initiate translation of the P₂₆ polypeptide.

The first AUG on mRNA transcribed from pGEMP23 is AUG₃₉₁. The largest species translated from this message is the P₂₃ polypeptide. There are no other AUG codons close enough on the mRNA to encode for a 23 kDa species and so AUG₃₉₁ is almost certainly responsible for initiation of the P₂₃ species. The second AUG on pGEMP23 transcribed mRNA is AUG₄₅₇ and this is probably responsible for initiation of the P₁₆ species, as there are no other AUG codons sufficiently close to allow production of a polypeptide of this size.

4.3: DISCUSSION.

4.3.1: *In vitro* expression of the PVM P protein.

Messenger RNA transcribed from PVM P gene cDNA directs the translation of a polypeptide with a molecular weight of 39 kDa and this confirms earlier work which assigned the 39 kDa PVM-infected cell polypeptide as being the PVM phosphoprotein (Cash, 1979; Ling and Pringle, 1989a). The predicted molecular weight of the PVM P protein based on the nucleotide sequence data presented here is 32,916 Daltons and so clearly the PVM P protein runs with aberrant mobility when analysed by SDS-PAGE. Similar discrepancies have been found between the predicted and observed molecular

weights of the HRSV P proteins (Johnson and Collins, 1990) and also for many other negative sense single stranded RNA viruses such as VSV (Gallione *et al.*, 1981). It has been suggested that the anomalous electrophoretic mobility of the P protein may be due to its characteristically high proportion of negatively charged residues. These amino acid residues are thought to bind poorly to SDS due to electrostatic repulsion, and thus the polypeptide becomes less mobile during electrophoresis. The PVM P₃₉ protein has been shown to be phosphorylated (Cash, 1979; Ling and Pringle, 1989a) although phosphorylation of the P protein is not thought to significantly contribute to its anomalous migration (Hsu and Kingsbury, 1982). It is not known at which residues the PVM P₃₉ protein is phosphorylated. The HRSV P protein is believed to be phosphorylated at serine residues at amino acid positions 116, 117, 119, 143, 156 and 161 (Navarro *et al.*, 1991). Interestingly, comparison between the PVM and HRSV P protein sequences reveals that the PVM P protein residues that are aligned to the HRSV amino acids 116, 119, and 143 are also serines and in addition the amino acids that align with HRSV residues 156 and 161 are threonines, which is also a potential phosphorylation site. Thus PVM P₃₉, like HRSV P may be phosphorylated at these sites.

4.3.2: *In vitro* expression of other P protein species.

Reticulocyte lysate translation of the mRNA transcribed from plasmid pGEMP gave rise not only to the 39 kDa P protein but also to minor species with observed molecular weights of 37, 26, 23 and 16 kDa. This observation supports hybrid arrested translation experiments which showed that PVM P gene cDNA was able to arrest the translation of not only P₃₉, but also P₃₇ and several smaller polypeptides which probably correspond to P₂₆, P₂₃ and P₁₆ (Chambers *et al.*, 1990a). By constructing transcription vectors containing truncated P gene cDNAs, the initiation codons that are

responsible for these P protein species have been mapped to AUG codons located at nucleotide positions 148, 301, 391 and 457 respectively. These are the second, third, fourth and fifth AUG codons respectively on the PVM P gene. In mapping these initiation codons the assumption was made that only AUG codons were capable of initiating translation. This assumption is almost certainly valid as the occurrence of non-AUG initiation codons is very unusual, although the C' polypeptide encoded by the P gene of SEN has been shown to be synthesised by initiation from an ACG codon (Gupta and Patwardhan, 1988; Curran and Kolakofsky, 1988a) and the C' polypeptide of PIV-3 is translated from a GUG codon (Boeck *et al.*, 1992). These non-AUG codons presumably function because they lie in excellent nucleotide context which partially offsets the absence of the A at position +1, or the G at position +2. Initiation of these PVM P protein species from an ACG codon is unlikely since all the mapped AUG codons are located on the P gene cDNA in such a place as to be capable of initiating synthesis of polypeptides having the correct molecular weight for the species to which each AUG was mapped. There are no ACG or GUG codons on the P gene mRNA which are in good context and are also in the vicinity of these mapped AUGs. To definitely demonstrate that the AUG codons at positions 10, 148, 301, 391 and 457 are responsible for these species the codons would have to be altered to non-initiating codons by site directed mutagenesis. A candidate for the polypeptide product of the P gene second ORF was not identified in the expression products of rabbit reticulocyte lysate programmed with mRNA transcribed from either pGEMP or pGEMP37. This was probably due to presence of globin which, in my hands, caused smearing of polypeptides in the 16-20 kDa region of the gel (see section 8.2).

4.3.3: Leaky scanning and the expression of the P protein species.

The way in which the P₃₇, P₂₆, P₂₃ and P₁₆ polypeptide species arise is probably through the process of leaky scanning (reviewed in Kozak, 1986a, 1986b,

1986c, 1989). According to this hypothesis a proportion of scanning 40S ribosomal subunits fail to initiate at a 5' proximal initiation codon which is surrounded by nucleotide sequence that renders it in a sub-optimal context. Having failed to initiate, the 40S subunits commence scanning for the next initiation codon. The actual proportion of ribosomes that fail to initiate at the sub-optimally situated initiation codon depends on just how poor the context is. When these ribosomal subunits reach the next potential initiation codon, they may initiate translation or if the context surrounding the codon is not favourable, they may once more resume scanning. The first AUG codon of the PVM P gene (AUG₁₀) is situated in a context that is considered to be highly favourable for initiation by virtue of similarities to the initiation consensus sequence which in the higher eukaryotes is CCACCAUGG (Kozak, 1986b). The G nucleotide at position +3 and the A nucleotide at position -4 are particularly critical for optimal context (Kozak, 1986a), and AUG₁₀ has both of these. The context of AUG₁₀ and the next four downstream AUG codons in the large ORF is displayed in figure 17. Since AUG₁₀ is the first potential initiation codon from the mRNA 5' terminus, a high proportion of ribosomal subunits would be expected to initiate at this position and only a few, if any, would be expected to continue scanning (Kozak, 1987). The context which surrounds the second AUG codon in-frame with the P gene long ORF located at nucleotide position 148 is not as favourable as that for AUG₁₀ since, whilst there is an A at position -3, there is a C at position +4 which renders this codon sub-optimal. However, this context is considered to be sufficiently favourable for initiation to occur by virtue of the purine at position -3. Since the context around AUG₁₄₈ is not optimally favourable for initiation, a proportion of those ribosomes that failed to initiate at AUG₁₀ may also fail to initiate at AUG₁₄₈ and

P protein ORF AUG codon number	P gene nucleotide	sequence surrounding AUG codon
1	10	UAA <u>A</u> AU AUG <u>G</u>
2	148	AUA <u>A</u> <u>C</u> AUG C
3	301	AGA <u>C</u> U AUG C
4	391	AGC <u>C</u> G AUG <u>G</u>
5	457	AGA <u>C</u> <u>C</u> AUG C
P2 ORF initiation codon		
AUG ₁₃₁		<u>C</u> ACU <u>C</u> AUG U

Figure 15. The nucleotide context that surrounds the first 5 initiation codons of the PVM P gene mRNA. For each of the P gene AUGs listed, the nucleotide identities with the optimal context motif (determined to be CCA/GCCAUGG; Kozak, 1986a) are shown in bold face and are underlined. The context of the first initiation codon of the P gene P2 ORF is also shown.

would be able to pass along to the next potential initiation codon. The downstream AUG initiation codons at nucleotide positions 301, 391 and 457 are also situated in context that is not optimal (either no G at position +4, or no A at position -3) but due to the presence either one of these critical nucleotides, may be sufficiently favourable for initiation to occur providing ribosomal subunits had progressed that far. The scenario described above would mean that due to the favourable nature of AUG₁₀ the major expression product of the P gene mRNA transcribed from pGEMP would be the full-length P protein and studies using rabbit reticulocyte lysate have shown that this is certainly the case. However when pGEMP transcribed P gene mRNA is translated *in vitro*, small amounts of the P₃₇, P₂₆, P₂₃ and P₁₆ polypeptides are clearly expressed. If these polypeptide species are expressed as a result of leaky scanning from their corresponding initiation codons (mapped as AUGs 148, 301, 391 and 457 respectively) then a number of ribosomal subunits must pass by the extremely favourable initiation codon AUG₁₀. Precisely what proportion of ribosome subunits that pass an initiation codon that is, by conventional rules, considered to be favourably situated, is not presently known, and so whether a sufficient number of ribosomes do pass AUG₁₀ in this case to support the expression of the P₃₇, P₂₆, P₂₃ and P₁₆ polypeptides is also unknown.

4.3.4: mRNA breakage and expression of the P protein species.

There are two other less likely but theoretically possible reasons that can account for the presence of these P protein related polypeptide species when full-length P gene mRNA is translated, and these must be considered. The first possibility is that *in vitro* transcribed full-length P gene mRNA may be degraded to create a series of fragments. mRNA breakage is thought to be responsible for activation of internal AUG codons in many systems (Kozak, 1989 and references therein). mRNA cleavage

would mean that the likelihood of translation being initiated from one of the normally internally located initiation codons becomes virtually assured due to its new position at the 5' end of the broken mRNA. This also means that initiation of translation from AUGs located further downstream of these newly positioned 5' proximal initiation codons also becomes more likely due to leaky scanning. For example, because of the favourable nature of AUG₁₀ for ribosomal initiation, a breakage anywhere in the mRNA between nucleotide 10 and the position of AUG₁₄₈ would probably result in translation from AUG₁₄₈ since it would now be the first AUG on the mRNA fragment, and because of the only moderate favourability of this initiation codon, initiation from the downstream AUG codons in accordance with the leaky scanning hypothesis described above could occur. The P₃₇, P₂₆, P₂₃ and P₁₆ polypeptides would be expressed in this way if only a small proportion of full-length P mRNAs became broken in this way. It is possible that the cleavage may occur at a variety of positions such that each of the in-frame AUG codons of the PVM P gene appears at the 5' end of a sub-set of the broken mRNA fragments. In this case, initiation could occur at each of the in-frame AUG codons, either by virtue of it being the first AUG on an mRNA fragment, or as a result of leaky scanning on an unfavourably situated upstream codon. The lack of a cap structure at the 5' end of each of the theoretically broken mRNAs would not prevent utilisation of these codons since reticulocyte lysate translation systems are known to be able to translate uncapped mRNAs.

The suggestion that leaky scanning gives rise to the smaller molecular weight P protein species is supported by the comparison of the polypeptide products that arise from pGEMP37 and pGEMP transcribed mRNA. Messenger RNA transcribed from pGEMP has the strong initiating codon AUG₁₀ in the 5' proximal position and consequently the full-length P protein is expressed in quantities vastly greater than the smaller P related polypeptides since this codon prevents the vast majority of 40S subunits from passing to downstream regions of the message. Messenger RNA transcribed from pGEMP37 has lost the strongly initiating AUG₁₀ and the less

favourable AUG₁₄₈ becomes the 5' proximal AUG codon. When this mRNA is translated the most abundant species is the P₃₇ polypeptide, but considerable quantities of the P₂₆, P₂₃ and P₁₆ species are also made. Due to their similarities the two messages are probably of similar stabilities and since they are transcribed under identical conditions are therefore expected to be equally intact. The increased abundance of the lesser molecular weight P protein species is therefore probably due to the context of codon AUG₁₄₈ which permits leaky scanning and not due to mRNA breakage.

Expression abundance from P gene mRNA of each of the P₃₉, P₃₇, P₂₆, P₂₃ and P₁₆ polypeptides in a reticulocyte lysate decreases in that order. This observation would be expected if initiation of the lower molecular weight P protein species was due to the process of leaky scanning, as the number of 40S ribosomal subunits scanning towards the AUG codons located further towards the 3' end of the message would be expected to drop as each successive AUG codon was passed. If mRNA breakage was responsible for expression of these species then in order to produce this profile of polypeptide products, there would have to be progressively fewer breakages of the mRNA at sites towards the 3' end of the mRNA. Such breakages would result in the creation of progressively fewer mRNA species that had the second, third, fourth, and fifth AUG codons respectively at the 5' end of the broken mRNAs which would thus allow these AUGs to be used for initiation. The chances of breakages occurring in just this specific way seems unlikely.

Another line of evidence arguing against mRNA breakage as the cause of low molecular weight P protein species production is that each new batch of mRNA transcribed from any of the transcription vectors containing truncated P gene cDNAs have given rise to a constant profile of expression products each time the mRNA is translated. If the broken mRNAs were generated by cleavage as a result of random hits by RNase attack then the proportions of the truncated P protein species would be expected to vary between batches of mRNA.

4.3.5: Proteolytic cleavage and expression of P protein species.

The other mechanism which could possibly give rise to the P Protein related species in the rabbit reticulocyte *in vitro* translation system is proteolytic cleavage. This mechanism is unlikely to be responsible for P protein species production for several reasons. The cleavage events would have to occur at sites on P₃₉ that precisely coincided with the position of corresponding AUG initiation codons on P gene mRNA that are responsible for expression of the lower molecular weight P protein species (P₃₇, P₂₆, P₂₃, P₁₆) and this is highly unlikely. No other P protein species other than P₃₉, P₃₇, P₂₆, P₂₃ and P₁₆ are expressed by translation of full-length P gene mRNA and if cleavage were responsible for generation of the lower molecular weight P protein related species, then cleavage events would also have to occur only at the sites and no others, and again this is unlikely. Furthermore, the absence of any other P protein related polypeptides other than P₃₉, P₃₇, P₂₆, P₂₃ and P₁₆ does not support the idea of proteolytic cleavage because cleavage of P₃₉ to yield these smaller polypeptides would also presumably generate low molecular weight polypeptide fragments.



CHAPTER 5

**BACTERIAL EXPRESSION OF THE PVM P PROTEIN AND PVM P
PROTEIN RELATED POLYPEPTIDES.**

5.1: THE *E. COLI* BL21 EXPRESSION SYSTEM.

E. coli strain BL21 contains an IPTG inducible chromosomal copy of the gene which expresses T7 RNA polymerase (Studier and Moffatt, 1986). This RNA polymerase is able to transcribe cDNA sequences placed under the control of the T7 promoter present on pGEM1 derived plasmids pGEMP and pGEMP37. Construction of these plasmids is detailed in sections 4.1 and 4.2.2. The transcription plasmids pGEMP and pGEMP37 were transformed into *E. coli* strain BL21 using the procedure detailed in section 2.5.2, and expressed as detailed in section 2.12.2. The expression products were subjected to SDS-PAGE and then analysed by Western blot analysis using monoclonal antibody 26/3/B5.

5.2: *E. COLI* BL21 EXPRESSION OF THE P PROTEIN AND P PROTEIN RELATED SPECIES.

In Western blot analysis using the expression products of *E. coli* BL21 transformed with pGEMP, monoclonal antibody 26/3/B5 (section 6.3) recognised one major species which was present in high abundance and four species with a lower molecular weight which were present at a lower abundance (figure 16). The major expression product of *E. coli* BL21 transformed with pGEMP that was recognised by antibody 26/3/B5, comigrated with the PVM P₁₀ protein translated in reticulocyte lysate using pGEMP transcribed mRNA. The four minor species recognised by antibody 26/3/B5 and expressed by *E. coli* BL21 under the direction of pGEMP migrated with molecular weights of 37 kDa, 26 kDa, 23 kDa and 16 kDa. Thus it would seem that *E. coli* BL21 is capable of recognising the P gene translation initiation signals (AUGs 10, 148, 301, 391 and 457) and termination signals responsible for the expression of the P protein and the four P protein related species. The expression

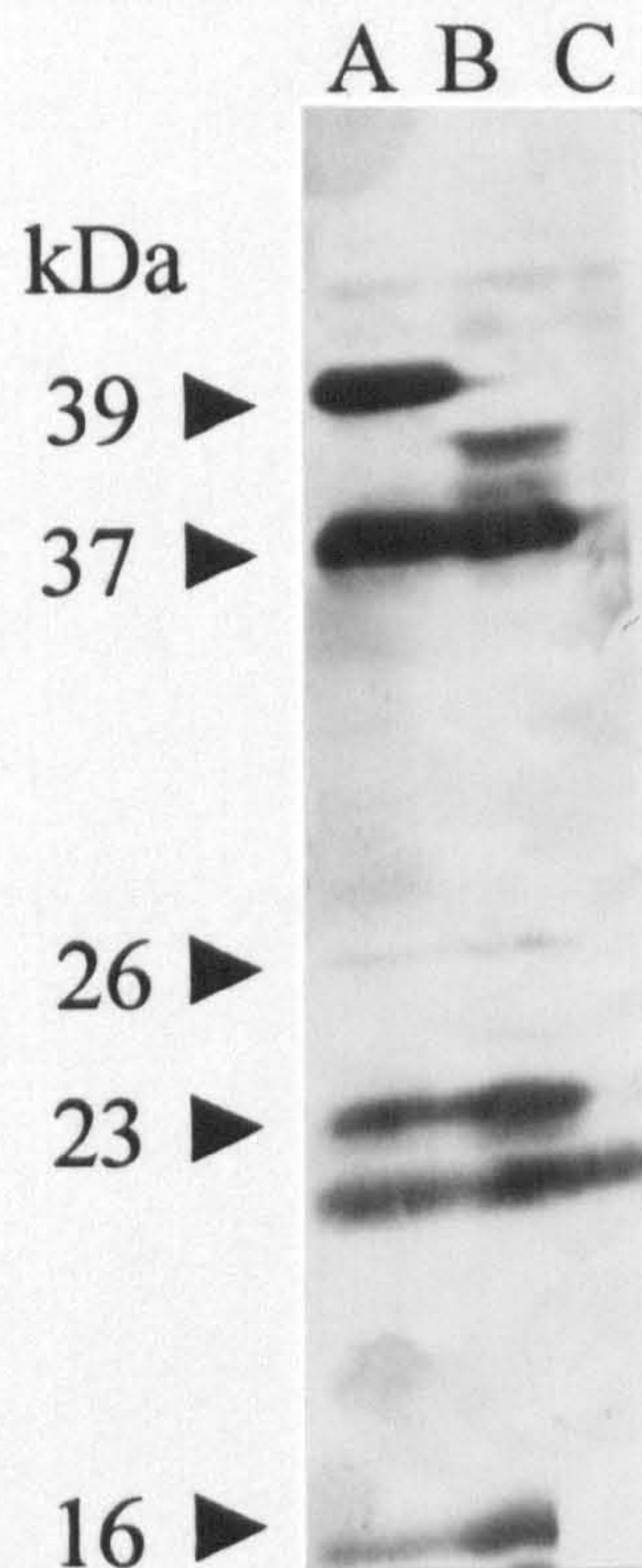


Figure 16. Expression of the 39 kDa PVM P protein and 37 kDa, 26 kDa, 23 kDa and 16 kDa P protein related polypeptides in *E. coli* BL21. *E. coli* BL21 was transformed with pGEMP (lane A) and pGEMP37 (lane B) and pGEM1 (lane C) and the expression products of IPTG-induced cultures were subjected to SDS-PAGE and then Western blotted using monoclonal antibody 26\3\B5. The positions of the P protein related polypeptides are indicated on the right and the positions of molecular weight markers are indicated on the left.

products of *E. coli* BL21 transformed with pGEMP37 were also analysed by SDS-PAGE and then Western blotted using monoclonal antibody 26/3/B5. In this system, pGEMP37 directed the synthesis of the P₃₇, P₂₆, P₂₃, and P₁₆ species (figure 16). By estimation of the signal intensity generated by antibody 26/3/B5 in Western blots, the quantity of each of the P protein species expressed by *E. coli* BL21 programmed with pGEMP37 was approximately equal to the quantity of their corresponding P protein species expressed by *E. coli* BL21 programmed with pGEMP. *E. coli* BL21 transformed with pGEMP26 expressed P₂₆, P₂₃ and P₁₆ in abundances that were also estimated as being equal to that of the corresponding P protein species expressed by pGEMP37 transformed *E. coli* BL21. Similarly, *E. coli* BL21 transformed with pGEMP23 expressed P₂₃ and P₁₆ at an abundance that was approximately equal to the level of these P protein species expressed in *E. coli* BL21 by either of pGEMP, pGEMP37 or pGEMP26.

5.3: DISCUSSION.

5.3.1: The P protein species expressed by *E. coli* BL21.

E. coli strain BL21 transformed with pGEMP was able to express the full-length PVM P protein and also the P-related polypeptides with molecular weights of 37, 26, 23 and 16 kDa. These polypeptides all precisely comigrated with their reticulocyte lysate expressed counterparts in SDS-PAGE suggesting that these P protein related polypeptides were being faithfully expressed and possibly post-translationally modified in the bacterial system. Thus T7 RNA polymerase expressed by *E. coli* BL21 correctly transcribed the P gene mRNA and the *E. coli* protein synthesis machinery was able to recognise the translational initiation and translational

termination signals present on the message. Transcription of the P gene mRNA would be expected to occur efficiently due to the presence of the T7 promoter sequence upstream of the P gene cDNA insert on pGEMP. Whether termination of transcription occurs at the 3' end of the P gene cDNA insert is not known as the pGEMP plasmid does not contain a transcription terminator which would prevent transcriptional read-through of sequences downstream. Expression of the correctly sized polypeptide from the T7 transcribed mRNA may therefore rely on the fidelity of translation. Clearly the ribosomes of *E. coli* BL21 are able to bind to the P gene mRNA and initiate translation. Inspection of the PVM P gene sequence reveals a short sequence that would be located at the very 5' terminal end of the subsequently transcribed message which partly resembles the Shine-Dalgarno consensus sequence (5'- AGGAGG -3') that makes up the prokaryotic ribosome binding site. This sequence, which is shown in figure 7, is 6 nucleotides away from the 5' proximal AUG codon of the P gene mRNA, and this spacing between Shine-Dalgarno sequence and initiating AUG codon is optimal according to the consensus spacing observed at *E. coli* mRNA initiation sites.

5.3.2: Leaky scanning and *E. coli* BL21 expression.

There is one way in which bacterial expression of the P₃₇, P₂₆, P₂₃ and P₁₆ polypeptides from full-length P gene cDNA supports the suggestion that these species, and their counterparts expressed in reticulocyte lysate, are translated by way of leaky scanning, and not because of mRNA breakage. With the exception of AUG₃₉₁, none of the potential initiation codons of the PVM P gene has a sequence that resembles the Shine-Dalgarno motif sufficiently closely to act as a ribosome binding site. Therefore if breakage of the P gene mRNA did occur after it was transcribed, it would probably not be translated because, with the exception of a fragment having AUG₃₉₁ at its 5' end, it would not possess a functional ribosome binding site. The potential ribosome binding

site located just upstream of AUG₃₉₁ almost perfectly matches the Shine-Dalgarno consensus, however this site may not function efficiently due to the observation that P₂₃ is poorly expressed from *E. coli* BL21 transformed with pGEMP23 which has AUG₃₉₁ as the 5' proximal AUG codon on the P gene cDNA.

Western blot analysis using an anti-P protein monoclonal antibody, showed that the only P related polypeptide expression products of *E. coli* BL21 transformed with either pGEMP or pGEMP37 were the P protein and P protein related species seen in the reticulocyte lysate expression experiments (figure 12) There was no evidence of any initiation events taking place on any non-AUG codons that are known to occur in prokaryotic systems.

CHAPTER 6

IDENTIFICATION OF MONOCLONAL ANTIBODIES SPECIFIC FOR THE PVM P GENE.

6.1: INTRODUCTION.

In order to effectively study the PVM P protein it was necessary to be able to unambiguously identify it amongst the multitude of polypeptide products expressed by PVM-infected BSC1 cells or the polypeptides expressed in bacterial culture. Specific identification would prevent any confusion that may arise from the presence of comigrating polypeptides, and in the case of the P protein this specificity would prevent confusion with the N protein which has a similar molecular weight and which is present in high abundance in PVM-infected BSC1 cells. The most effective tool with which the P protein could be specifically identified was with anti-P protein monoclonal antibodies.

Hybridoma cell lines secreting PVM-specific antibodies were produced and screened by Dr. R Ling as described in Ling and Pringle (1989a). By screening these monoclonal antibodies against PVM-infected BSC1 cell proteins in Western blot analysis, the specificity of four monoclonal antibodies was established. One of these antibodies was found to be specific for a PVM 39 kDa polypeptide and another was found to be specific towards either this 39 kDa species or the PVM N protein which has an apparent molecular weight of 42 kDa and which was not well resolved using the gel system used by these authors. The monoclonal antibody specific for the 39 kDa PVM polypeptide, designated 26/4/C4, also cross-reacted with the HRSV P protein (Ling and Pringle, 1989a). For this reason it was proposed that 26/4/C4 was specific for the PVM P protein. The second antibody, 26/1/A2 did not cross-react with a HRSV polypeptide and the identity of the PVM polypeptide for which it is specific was ambiguous.

In order to be able to more fully characterise the expression products of the PVM P gene, approximately two hundred unscreened hybridoma cell lines prepared by Dr. R. Ling were screened for ability to recognise PVM P protein in the hope of identifying further P protein specific monoclonal antibodies

6.2: MONOCLONAL ANTIBODY SCREENING PROCEDURE.

The tissue culture fluids collected by Dr. R. Ling from antibody-expressing hybridoma cell lines were used in Western blot analysis against PVM-infected BSC1 cell material which had been subjected to SDS-PAGE and transferred to PVDF membrane. The PVM-infected and mock-infected BSC1 cell material immobilised onto the membrane was cut into thin strips 4 mm wide and one strip of each PVM-infected and mock-infected BSC1 cell material was probed with each antibody (figure 17). A total of 20 monoclonal antibodies recognised a band in the region of the blot corresponding to the position of the putative PVM P protein. However, as was mentioned above, the PVM N protein migrates close to the position of the PVM P protein and so to unequivocally determine their specificity these 20 monoclonal antibodies were used in further Western blot analysis using the PVM P protein and amino-terminally truncated P protein species expressed in *E. coli* BL21 transformed with pGEMP39 and pGEMP37. As described in section 4.2, pPEMP39 directs the synthesis of not only the full-length P protein but also the P₃₇, P₂₆, P₂₃ and P₁₆ species. The quantity in which this system expressed the P₁₆ species was not sufficient for reliable identification of this polypeptide to be made in Western blot analysis. In a similar way, pGEMP37 directs the synthesis of the P₃₇, P₂₆, P₂₃ and P₁₆ polypeptide species. The expression of the P protein related polypeptides is discussed in section 4.3.2.

6.3: EPITOPE MAPPING OF PVM P PROTEIN SPECIFIC MONOCLONAL ANTIBODIES.

Because the P₃₇, P₂₆, P₂₃, and P₁₆ polypeptides all possess the carboxyl-terminal region of the PVM P protein but various amounts of the P protein amino-

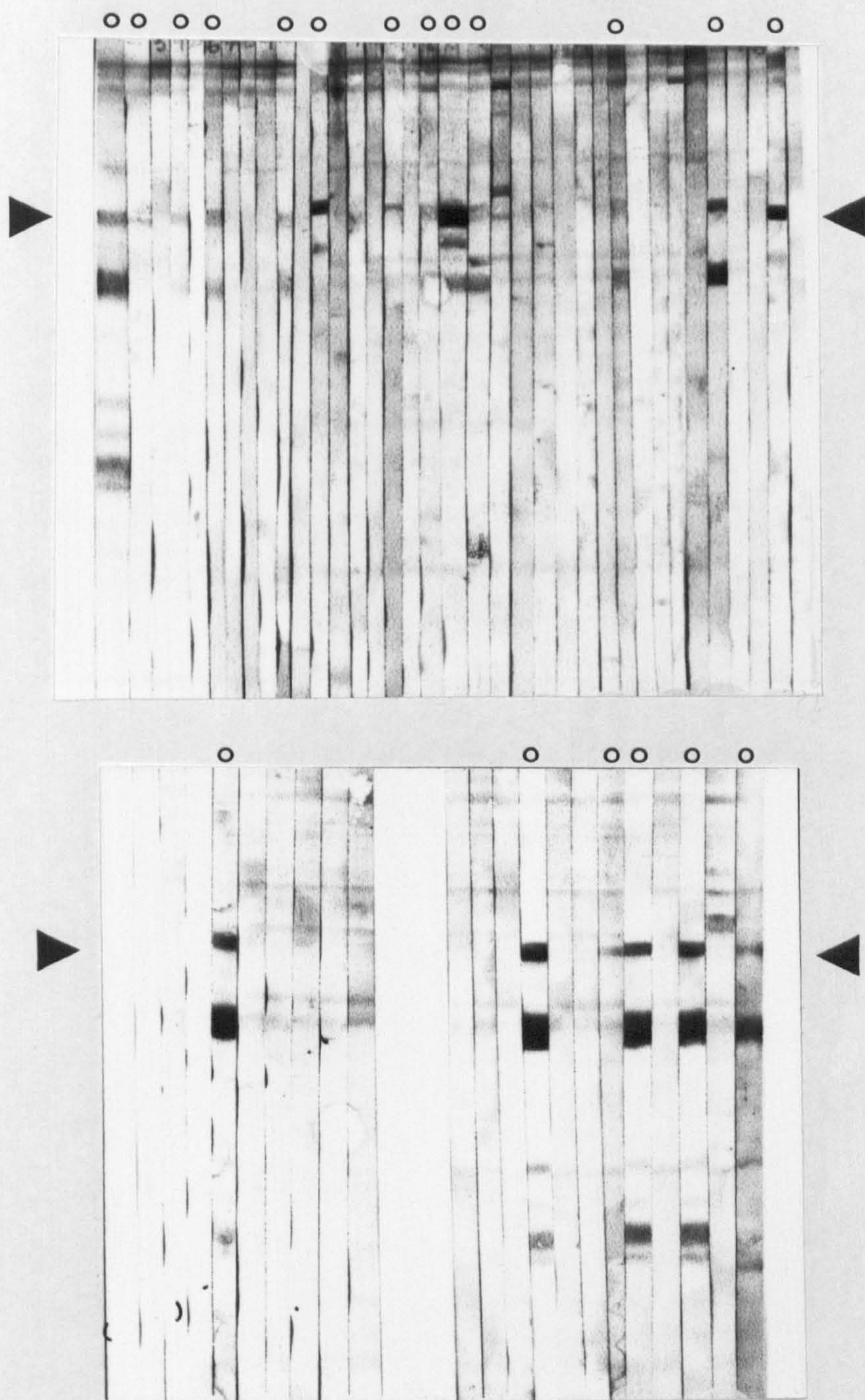


Figure 17. Screening of monoclonal antibodies by Western blot analysis with PVM-infected BSC1 cells. The position of the PVM P39 protein is indicated. The monoclonal antibodies (o) that were thought to be specific for the P protein were used in further Western blot analysis using bacterially expressed P protein.

terminus, recognition of all or just some of these species by a monoclonal antibody reveals the approximate location of the P protein epitope that the monoclonal antibody is able to recognise. Each of the 20 monoclonal antibodies identified as being specific for the P protein were used in Western blot analysis against P protein expressed in PVM-infected BSC1 cells and in bacterial cultures. The set of P protein species that these antibodies recognised fell into one of three groups. Figure 18 displays the results of the Western blot analysis by listing which of the four P protein species originating from both PVM-infected BSC1 cells and from bacterial expression that the anti-P protein monoclonal antibodies were capable of recognising. This figure also displays the region of the PVM P protein that the epitopes recognised by these antibodies can be mapped to. A Western blot representative of the reaction pattern of two of the three groups is shown in figure 19. The monoclonal antibody 26/4/C4 (Ling and Pringle, 1989a) which constitutes the third group is also included in this table.

The position of the epitopes for these antibodies do not appear to be evenly spread throughout the length of the P protein. One of the 21 antibodies listed in figure 18 recognised an epitope in the extreme amino-terminal 47 amino acids of the P protein, thirteen antibodies recognised a site within amino acids 48 to 98, and the remaining seven recognised a site somewhere in the carboxyl-terminal 157 amino acids. The region spanning amino acids 48 to 98 is the location for the epitopes of thirteen antibodies and this figure is four times that which would be expected if all sites of the P protein were equally likely to be epitope sites. Whilst it is possible that the low number of antibodies used in this study means the uneven distribution of epitope sites may not be statistically significant, it may also reveal that amino acids 48 to 98 of the native PVM P protein that was used to immunise the mice for the production of antibodies was exposed on the surface of the molecule and thus was in a position to elicit an immune response. Arguing against this suggestion, the hydropathy plot of the PVM P protein (section 3.4.3) reveals that this region of the

PVM P protein is predominantly hydrophobic which suggests that this region will not be exposed. More detailed analysis of the epitopes recognised by these antibodies would be beneficial as it would enable more detailed investigation of the lower molecular weight PVM P protein species. This would be possible by bacterially expressing smaller regions of the PVM P protein and subjecting these P protein polypeptides to Western blot analysis using the above monoclonal antibodies.

(A)	MONOCLONAL ANTIBODY	P-PROTEIN SPECIES RECOGNISED	REGION MAPPED ON P PROTEIN
	26/4/C4*	39	A
	26/2/F8	39, 37	B
	26/1/E4	39, 37	B
	24/4/A4	39, 37	B
	26/3/F11	39, 37	B
	26/1/A2	39, 37	B
	26/2/E9	37, 37	B
	26/4/E6	39, 37	B
	26/1/D7	39, 37	B
	26/4/D10	39, 37	B
	26/3/G4	39, 37	B
	25/1/C10	39, 37	B
	24/4/C9	39, 37	B
	24/3/A7	39, 37	B
	26/1/D12	39, 37	B
	24/1/B7	39, 37, 26, 23	C
	26/3/E2	39, 37, 26, 23	C
	26/4/C10	39, 37, 26, 23	C
	26/2/C6	39, 37, 26, 23	C
	26/1/F6	39, 37, 26, 23	C
	26/3/B5	39, 37, 26, 23	C

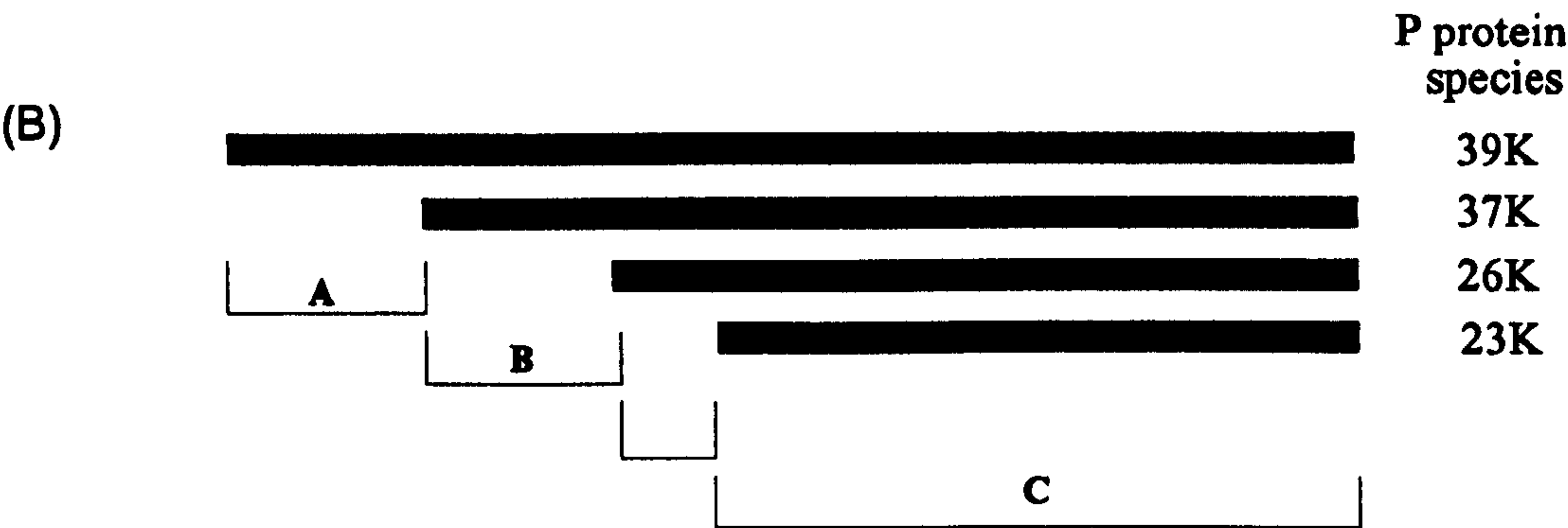


Figure 18. (A) A listing of the 21 anti-P protein monoclonal antibodies that were characterised. *Antibody 26/4/C4 was characterised by Dr R. Ling (Ling and Pringle, 1989a). The ability of the monoclonal antibodies to recognise bacterially expressed P₃₉, P₃₇, P₂₆ and P₂₃ P-related proteins is shown. The ability to recognise some or all of these polypeptides identified the approximate position of the P protein epitope that the monoclonal antibody was able to recognise. (B). A diagrammatical representation of the PVM P protein showing the three regions to which the epitopes recognised by the 21 monoclonal antibodies could be mapped to. For example, antibody 26/3/B5 recognised P₃₉, P₃₇, P₂₆ and P₂₃, therefore it must recognise an epitope in region C since only this region is common to all four P protein species. The recognition of the 16 kDa P related polypeptide is not recorded here as it was expressed at low abundance and was not reliably identified.

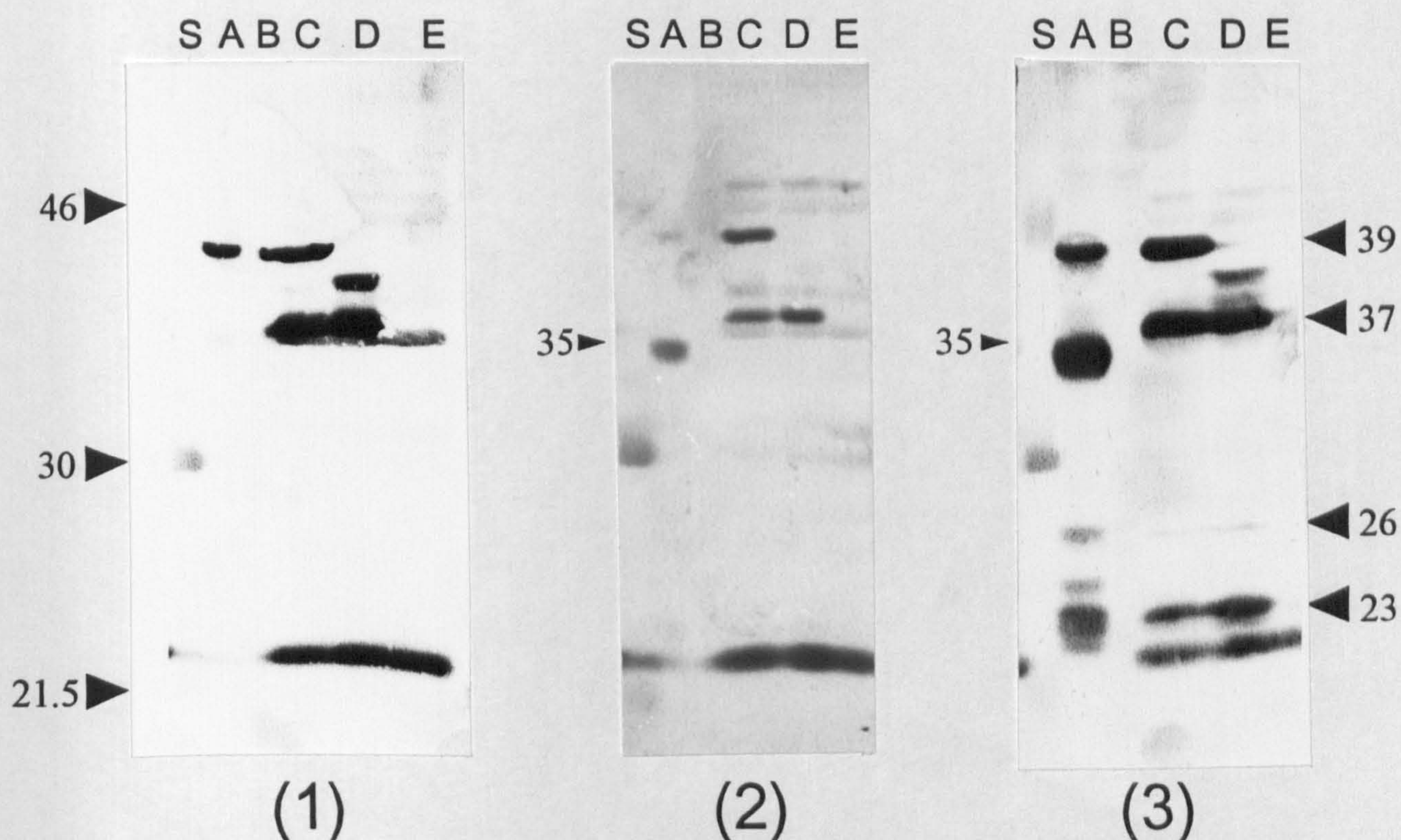


Figure 19. Mapping the regions of the P protein to which anti-P protein monoclonal antibodies bind. Three monoclonal antibodies specific for the PVM P protein were used in Western blot analysis with PVM-infected cell material, mock-infected cell material and also the expression products of *E. coli* BL21 that had been transformed with transcription plasmids pGEMP, pGEMP37 and pGEM1. Plasmid pGEMP is able to direct the expression of the full-length P protein and also the 37 kDa, 26 kDa, 23 kDa and 16 kDa P protein related polypeptides, whereas pGEMP37 directs the expression of the four lower molecular weight species. Plasmid pGEM1 does not direct the expression of any P protein related polypeptides and was used as a control. The lanes are assigned as follows. Lane A: PVM-infected cell material. Lane B: Mock-infected cell material. Lane C: IPTG-induced *E. coli* BL21 transformed with pGEMP. Lane D: IPTG-induced *E. coli* BL21 transformed with pGEMP37. Lane E: IPTG-induced *E. coli* BL21 transformed with pGEM1. Depending on the reactivity with the P protein species expressed both in PVM-infected cells and expressed in bacteria, a region of the P protein that is recognised by the antibodies can be determined (see section 6.3 for explanation). The monoclonal antibodies used in the blots shown above are (1) 26/3/G4. (2) 26/1/D7. (3) 26/3/E2. The position of the 39 kDa P protein and the 37 kDa, 26 kDa and 23 kDa polypeptides are indicated on the right. The 16 kDa P protein species is expressed in low abundance and was not detected by any of these antibodies. The position of the 35 kDa P protein species expressed in PVM-infected cells is shown and the position of molecular weight markers (lane S) is indicated on the left.

CHAPTER 7

**THE PVM P PROTEIN SPECIES EXPRESSED IN PVM-INFECTED BSC1
CELLS.**

7.1: WESTERN BLOT ANALYSIS OF PVM-INFECTED BSC1 CELL MATERIAL.

In Western blot analysis using PVM-infected BSC1 cells, the anti-P protein monoclonal antibody 26/3/B5 identified eleven polypeptide species with estimated molecular weights of between 39 kDa and 14 kDa (figure 20). None of these species were seen in mock-infected BSC1 cells. These eleven species were routinely identified by antibody 26/3/B5 using several different batches of PVM-infected BSC1 cell material. These polypeptide species were also routinely identified by several other anti-P protein monoclonal antibodies described in section 6.3. As can be seen from figure 20 the P protein related polypeptides recognised by monoclonal antibody 26/3/B5 were not all of equal abundancies. The most abundant species were the P₃₉ polypeptide and also a polypeptide with a molecular weight of 35 kDa, which is discussed in section 7.3.2.

7.2: COMPARISON OF P PROTEIN SPECIES EXPRESSED IN PVM-INFECTED BSC1 CELLS WITH P PROTEIN SPECIES EXPRESSED IN *E. COLI* BL21.

The expression products of *E. coli* BL21 transformed with pGEMP were electrophoresed alongside PVM-infected and mock-infected BSC1 cell species and then Western blotted using monoclonal antibody 26/3/B5 (figure 21). This antibody recognised eleven P protein related species expressed in PVM-infected BSC1 cells and also five P protein species expressed in *E. coli* BL21 which have molecular weights of 39 kDa, 37 kDa, 26 kDa, 23 kDa and 16 kDa. The P₃₉ protein expressed in PVM-infected BSC1 cells comigrated exactly with the bacterially expressed 39 kDa P protein. In addition, the 26 kDa and 23 kDa P protein related polypeptides expressed from *E. coli* BL21, comigrated exactly with species expressed in PVM-infected BSC1 cells. Previous work described in sections 4.2.3 and 5.2 showed that the 26 kDa and

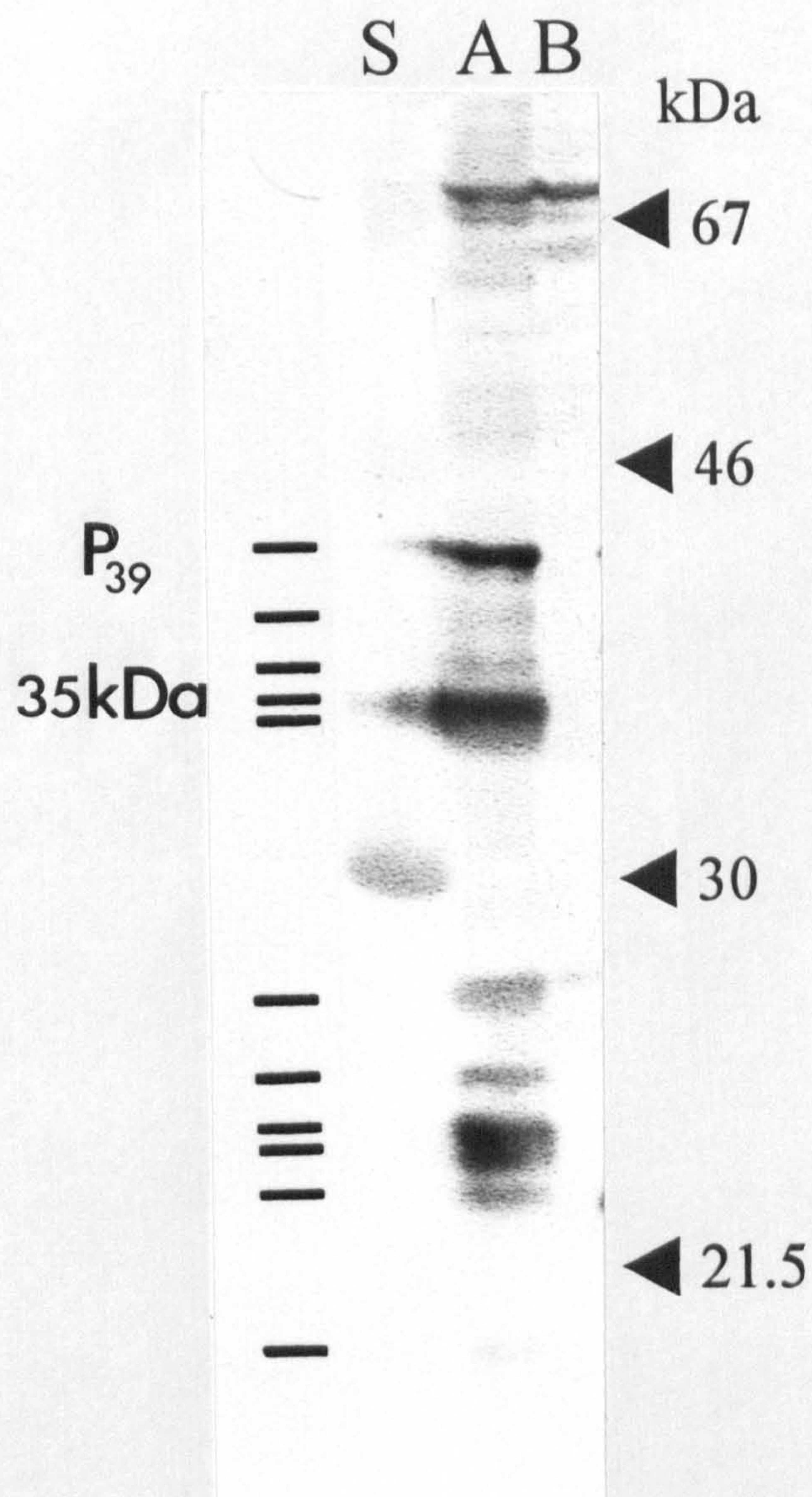


Figure 20. In addition to the P protein, several P protein related polypeptides are synthesised in PVM-infected BSC1 cells. PVM-infected cell material (lane A) and mock-infected cell material (lane B) was subjected to SDS-PAGE and Western blotted using anti-P protein monoclonal antibody 26/3/B5 (section 6.3). The positions of the P protein and the P protein specific polypeptides are indicated with horizontal bars on the left. The positions of molecular weight markers (lane S) are indicated on the right.

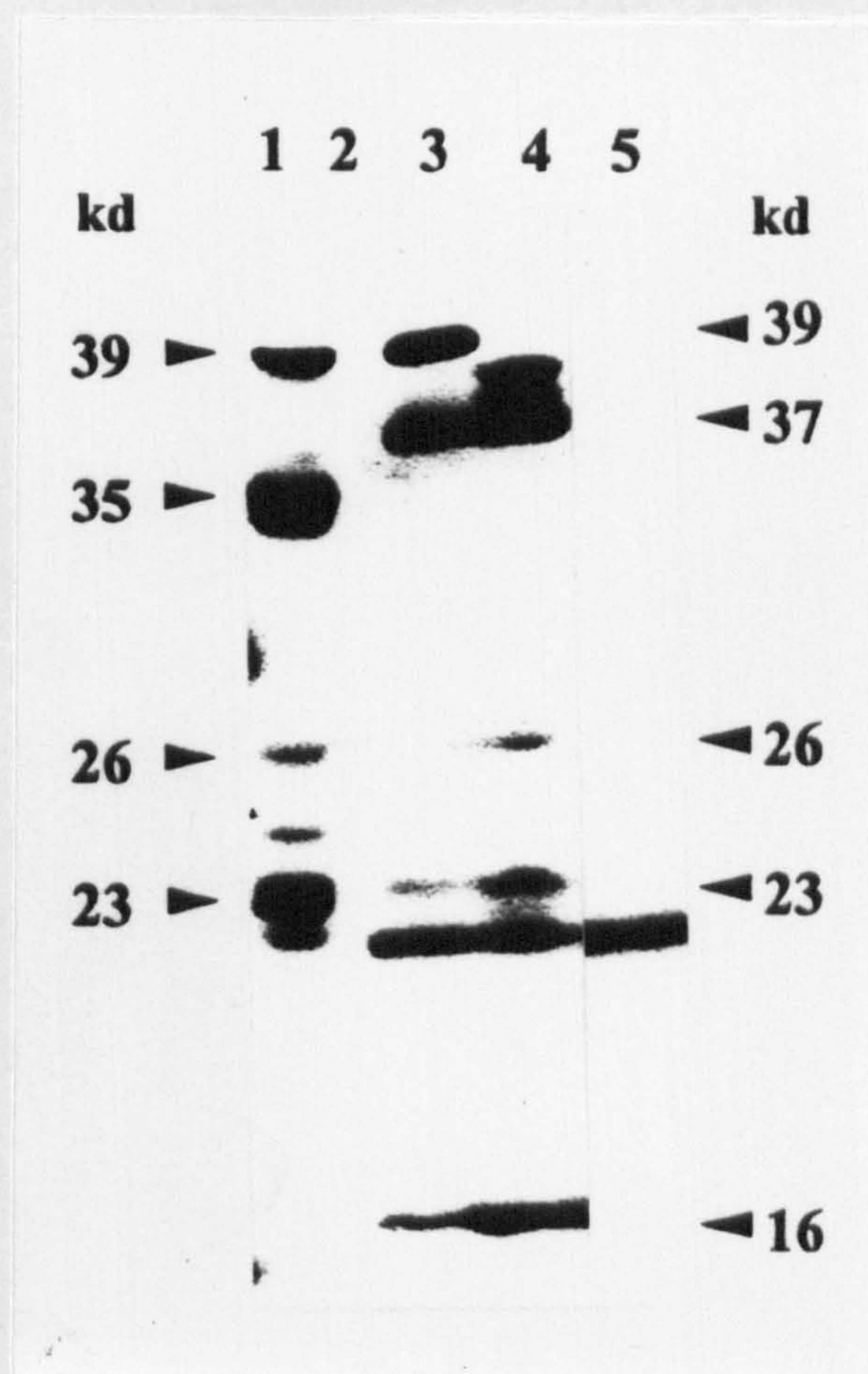


Figure 21. The PVM P gene long open reading frame encodes at least 3 products *in vivo*. Transcription plasmids pGEMP and pGEMP37 were transformed into *E. coli* BL21 which is capable of expressing the proteins encoded for by these plasmids. *E. coli* BL21 expression products were subjected to SDS-PAGE along with PVM-infected and mock-infected BSC1 cell material and then Western blotted using monoclonal antibody 26\3\B5 (section 6.3). Lane 1: PVM-infected BSC1 cell material. Lane 2: Mock-infected BSC1 cell material. Lane 3: Expression products of *E. coli* BL21 transformed with pGEMP. Lane 4: Expression products of *E. coli* BL21 transformed with pGEMP37. Lane 5: Expression products of *E. coli* BL21 transformed with pGEM1. The molecular weights of the marked polypeptides are indicated and are expressed in kilodaltons.

23 kDa *E. coli* BL21 expressed P protein related polypeptides are expressed from initiation codons AUG₃₀₁ and AUG₃₉₁ of the P gene long ORF. Recognition by monoclonal antibody 26/3/B5 identifies the PVM-infected BSC1 cell 26 kDa and 23 kDa proteins as being expressed from the PVM P gene. The expression of the *E. coli* BL21 26 kDa and 23 kDa species only in the cultures that contained pGEMP and not pGEM1 identifies these polypeptides as being products of the P gene. The observation that the bacterially expressed 26 kDa and 23 kDa P protein species comigrate with the PVM-infected BSC1 cell 26 kDa and 23 kDa P protein species suggests that the PVM-infected BSC1 cell polypeptides represent P protein species possessing identical amino acid sequence to the bacterially expressed species. Apart from these three P protein related polypeptides with molecular weights of 39 kDa, 26 kDa and 23 kDa, none of the other eight PVM-infected BSC1 cell species comigrated with an *E. coli* BL21 expressed P protein related species.

7.3: DISCUSSION.

7.3.1: The 26 kDa and 23 kDa P protein species expressed in PVM-infected BSC1 cells.

The *E. coli* BL21 expressed 39 kDa, 26 kDa and 23 kDa P protein related polypeptides each comigrated with a P protein related species expressed from PVM-infected BSC1 cells. The 39 kDa *E. coli* BL21 expressed P protein related species corresponds to the full-length P₃₉ protein expressed in PVM-infected BSC1 cells. The *E. coli* BL21 expressed species with molecular weights of 26 kDa and 23 kDa comigrate with the 26 kDa species and 23 kDa species expressed in PVM-infected BSC1 cells. These *E. coli* expressed 26 kDa and 23 kDa species are translated by

initiation on AUG₃₀₁ and AUG₃₉₁ of the PVM P gene respectively (sections 4.2.3 and 5.2) and it is likely that the 26 kDa and 23 kDa polypeptides seen in PVM-infected BSC1 cells are also expressed from these respective AUG codons. The possibility that these two PVM-infected BSC1 cell species comigrate with the bacterially expressed species and yet represent different regions of the PVM P protein is unlikely. If the 23 kDa and 26 kDa species expressed in PVM-infected BSC1 cells could be shown to be amino-terminally truncated version of the full-length P protein, then this would support the proposition that these polypeptides represent the same amino acid sequence as the bacterially expressed 26 kDa and 23 kDa polypeptides. A clue as to which region of the P protein the PVM-infected BSC1 cell expressed 26 kDa and 23 kDa species represent comes from observations noted whilst identifying the position on the P protein where the anti-P protein monoclonal antibody epitopes were situated (section 6.3). When a particular monoclonal antibody recognised the *E. coli* BL21 expressed P protein species, the PVM-infected BSC1 cell protein species of identical molecular weight was also always recognised. This observation was true for all the P protein specific monoclonal antibodies analysed. If the PVM-infected BSC1 26 kDa and 23 kDa species did not represent the same region of the full-length PVM P protein as their comigrating *E. coli* BL21 expressed counterparts then instances would arise where the monoclonal antibody epitope was missing from either the PVM-infected BSC1 cell or the bacterially expressed species and so the comigrating pair would not both be recognised by the same monoclonal antibody. Neither the 26 kDa or the 23 kDa species can be derived from full-length P protein by removal solely of carboxyl-terminal sequence because they would be recognised by an antibody that recognised an amino-terminal epitope such as 26/4/C4 (Ling and Pringle, 1989a), or any of the antibodies that only recognise the 39 kDa and 37 kDa P protein species (section 6.3). This suggests that the 26 kDa and 23 kDa species seen in PVM-infected BSC1 cells do not possess the P protein amino-terminus, and it is probable that they arise respectively by way of internal initiation on AUG₃₀₁ and AUG₃₉₁ of the PVM P gene.

7.3.2: Other P protein related species expressed in PVM-infected BSC1 cells.

Out of the eleven P protein related polypeptides identified by monoclonal antibody 26/3/B5, the identity of eight is still unknown. One of these remaining eight uncharacterised polypeptides is a highly abundant species having a molecular weight of 35 kDa (figure 20). The precise identity of this polypeptide has not been established, however, there are several possibilities as to how this species might arise.

The 35 kDa species may be generated by cleavage from the full-length P protein. The cleavage may arise either as a consequence of the procedure used to harvest PVM-infected BSC1 cell material, whereby the full-length P protein could be chemically degraded by cleavage at either the amino or carboxyl-termini to remove a 4 kDa fragment, and so yield a 35 kDa polypeptide, or alternatively by way of proteolytic cleavage by a host BSC1 cell protease. In argument against the suggestion that the 35 kDa polypeptide is an artefact of the cell harvesting procedure is the observation that the full-length P protein expressed from *E. coli* BL21, which is harvested in an identical procedure to that of PVM-infected BSC1 cell material, is not cleaved or degraded to form a 35 kDa polypeptide. This observation that the 35 kDa species is not represented in the expression products of *E. coli* BL21 transformed with pGEMP also suggests that the agent responsible for cleaving full-length P protein may be present in BSC1 cells but not in *E. coli* BL21, and such an agent may be a BSC1 protease.

Alternatively, since a polypeptide with a molecular weight of 37 kDa which could correspond to the P₃₇ protein species expressed *in vitro* from P gene initiation codon AUG₁₄₈ was not identified by monoclonal antibody 26/3/B5 in PVM-infected BSC1 cell material, this abundant 35 kDa species may represent a post-translationally modified form of the *in vitro* expressed P₃₇ species. According to this scheme, AUG₁₄₈

would be utilised to generate the 37 kDa polypeptide and this would then be modified to generate the 35 kDa species.

Whilst mapping the sites on the PVM P protein where anti-P protein monoclonal antibody epitopes were located using P protein polypeptides expressed from both PVM-infected BSC1 cells and *E. coli* BL21 transformed with pGEMP, it was noticed that whilst some of the anti-P protein monoclonal antibodies were able to recognise the 39 kDa, 37 kDa and 35 kDa P protein species (such as antibody 26/1/D7, figure 19 [1]) some monoclonal antibodies only recognised the 39 kDa and 37 kDa species (such as antibody 26/3/G4, figure 19 [2]). This observation suggests that the PVM-infected BSC1 cell 35 kDa P protein related polypeptide is amino-terminally truncated to a greater extent than the 37 kDa protein which does not possess the amino-terminal 40 amino acids. Pulse-chase labelling of PVM-infected BSC1 cells may be able to establish a relationship between the 35 kDa polypeptide and either of the full-length P protein or the 37 kDa P species.

The remaining seven P protein specific polypeptides identified in PVM-infected BSC1 cells by monoclonal antibody 26/3/B5 are of unknown origin. They are unlikely to be the unmodified products of internal initiation events due to their molecular weights not corresponding to the P protein related species expressed *in vitro* from the internal initiation codons of the PVM P gene. They may represent P protein breakdown products, and since none of these seven polypeptides are seen in the expression products of *E. coli* BL21 transformed with pGEMP, they may be cleaved as a consequence of the PVM-infected BSC1 cell harvesting procedure or they may be cleaved by a BSC1 host cell protease. Some of the polypeptides may represent P protein species that have been phosphorylated to various extents as suggested for P protein related polypeptides seen in Newcastle disease virus infected cells (McGinnes *et al.*, 1988). The P genes of *Paramyxoviridae* family members are unusual in that many are able to direct the expression of several polypeptide products

and they are able to utilise more than one reading frame in order to do this. The P gene of SEN has been particularly well studied and this gene is now known to direct the synthesis of at least 8 polypeptides by using all three of its reading frames. A review of the data accumulated on this gene is given in the introduction, section 1.4.3.1. The 26 kDa and 23 kDa P protein species seen in PVM-infected BSC1 cells are analogous to the SEN X protein in that they are derived from the same ORF as their respective P proteins. The SEN X protein has not been fully characterised and the exact way in which it arises has yet to be established, however it appears to be initiated from either of two internal codons located at nucleotide positions 1505 or 1523 of the SEN P gene mRNA (Curran and Kolakofsky, 1988b). There are other examples whereby the large ORF of the P gene of a negative sense single stranded RNA virus gives rise to more than one species *in vivo*. A polypeptide representing the carboxyl-terminal 93 amino acids of the HRSV P protein was detected in very small quantities in HRSV-infected cells (Caravokyri *et al.*, 1992). This polypeptide, which had an observed molecular weight of 16 kDa, was believed to be initiated by leaky scanning, on an AUG codon (codon 148) which is the second AUG initiation codon in frame with the P ORF. It was suggested that the 16 kDa polypeptide may interact with the basic HRSV N and 22K proteins since it consists mainly of the acidic domains of the P protein. Another example is in the expression of a 7 kDa polypeptide from the vesicular stomatitis virus P protein ORF (Herman, 1986; 1987). This species is thought to be expressed from the 3' proximal AUG of the P protein mRNA and represents the carboxyl-terminus of the VSV P protein. No function has been assigned to this polypeptide. The P gene of NDV is believed to utilise the P protein ORF to express polypeptides with molecular weights of 38 kDa and 29 kDa (McGinnes *et al.*, 1988) although this interpretation has been questioned (Samson *et al.*, 1991).

CHAPTER 8

IDENTIFICATION OF THE PVM P GENE SECOND ORF PROTEIN PRODUCT (P2) IN PVM-INFECTED BSC1 CELLS.

8.1: EXPRESSION OF P2 IN A WHEAT GERM LYSATE SYSTEM.

In order to visualise the second open reading frame polypeptide product, mRNA transcribed from pGEMP37 was translated in a wheat germ lysate in the presence of ^{35}S methionine. Radiolabelled translation products were subjected to SDS-PAGE and an autoradiograph of the gel is shown in figure 22. A polypeptide with an observed molecular weight of approximately 20 kDa was identified in the radiolabelled translation products, and this was the most abundant polypeptide species expressed by the mRNA transcribed from pGEMP37. This 20 kDa polypeptide is not present in the translation products of mRNA transcribed from either pGEMP26 or pGEMP23, and therefore it must be initiated from an AUG codon situated upstream of AUG₃₀₁. There is only one AUG codon upstream of AUG₃₀₁ capable of initiating translation of a polypeptide of approximately 20 kDa and that is AUG₁₃₁ which is the first initiation codon of the second ORF. This AUG codon is the first initiation codon of the mRNA transcribed from pGEMP37. The 20 kDa polypeptide is the most abundant species expressed in wheat germ lysate from pGEMP37 transcribed mRNA and this supports the suggestion that this species is initiated from the first AUG codon on the mRNA. The putative second ORF polypeptide migrates with an apparent molecular weight of 20 kDa which is larger than the predicted molecular weight of 15,962 Da calculated from the second ORF polypeptide amino acid sequence.

8.2: TRANSLATIONAL INITIATION OF THE P2 POLYPEPTIDE *IN VITRO*.

The initiation codon that is responsible for expression of the second ORF polypeptide is the first AUG on the message transcribed from pGEMP37. This codon, which is located at nucleotides 131 to 133 of the P gene mRNA (figure 7), is situated

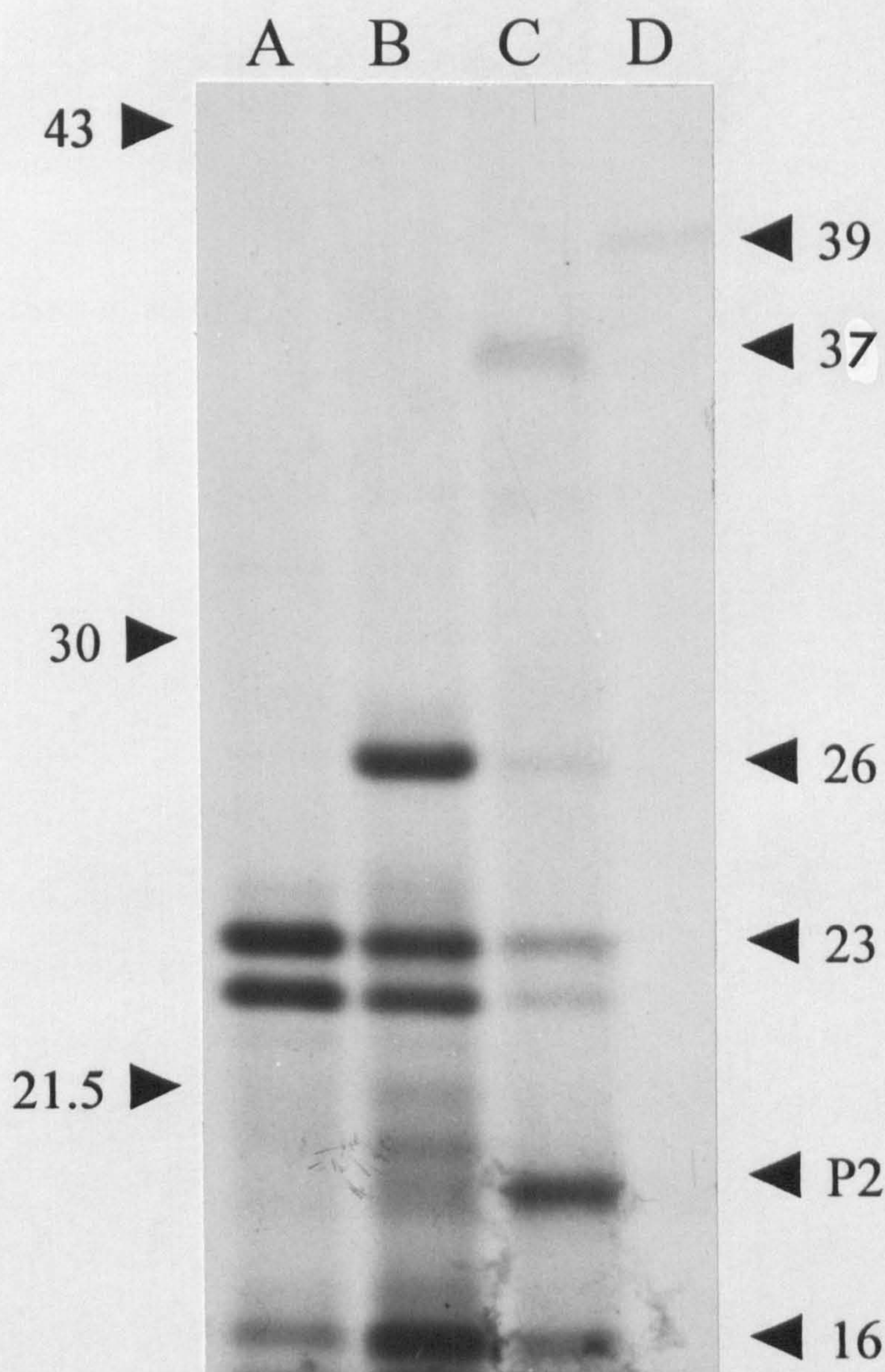


Figure 22. *In vitro* expression of the polypeptide product (P2) of the PVM P gene second open reading frame. A wheat germ lysate was programmed with mRNA transcribed *in vitro* from T7 transcription plasmids which contained P gene cDNA inserts. The radioactively labelled translation products were subjected to SDS-PAGE and the gel was autoradiographed in order to visualise the polypeptides. The plasmids used to transcribe the mRNA are as follows. pGEMP23 (Lane A), pGEMP26 (lane B), pGEMP37 (lane C), pGEMP (lane D). The P2 polypeptide is visible only in the expression products of wheat germ lysate programmed with mRNA transcribed from pGEMP37. The first initiation codon of the P2 ORF is the 3' proximal AUG codon on pGEMP37 transcribed mRNA. The ATG corresponding to this initiation codon is removed in the construction of pGEMP26 and pGEMP23. The position of molecular weight markers is indicated on the left.

in a context that is considered to be relatively unfavourable for initiating translation since it has neither a G at position +4 nor a purine at position -3 and it has only 2 of the cytosine residues that are ideally located at positions -5, -4, -2, and -1 (Figure 15; Kozak, 1986a; 1986b). Presumably the positioning of AUG₁₃₁ at the 5' end of the mRNA is sufficient to ensure that initiation will occur at this codon (Kozak, 1987). The mRNA transcribed from pGEMP37 expresses several other polypeptides, namely P₃₇, P₂₆, P₂₃ and P₁₆, in addition to the second ORF polypeptide, P2, and this probably results from a poor frequency of initiation from AUG₁₃₁ which allows scanning ribosomes access to downstream AUG codons. The expression of polypeptides that are initiated from these downstream AUG codons is interesting as they are not in the same ORF as the first initiation codon, AUG₁₃₁. This observation suggests that the scanning ribosomal subunits are not uniquely phased, either when they enter the mRNA species or when they encounter the first AUG codon. It has been suggested that the 40S ribosomal subunits may be nudged into the correct phasing by the GCC or ACC nucleotide sequence that immediately precedes the majority of AUG codons (Kozak, 1989). Perhaps in the case of the pGEMP37 transcribed mRNA, a large proportion of 40S subunits pass by AUG₁₃₁ because of the poor nucleotide context, and eventually encounter AUG₁₄₈ which, by virtue of the immediately preceding nucleotides, is able to nudge the incoming ribosomes into the correct phase for expression of the 37 kDa polypeptide.

The second ORF polypeptide was not seen when reticulocyte lysate was programmed with mRNA transcribed from pGEMP37 and this was probably due to the presence of large quantities of globin which causes a smearing of polypeptide species that run in the lower region of the gel where the second ORF protein product migrates. Radiolabelled second ORF polypeptide product was not identified by SDS-PAGE when wheat germ lysate was programmed with mRNA transcribed from pGEMP. The context of AUG₁₀ is extremely favourable for initiation to occur at this point (section

4.3.3; figure 15) and so only a small proportion of 40S ribosomal subunits would be expected to pass by AUG₁₀ and to scan towards downstream initiation codons. The context of AUG₁₃₁ is poor and so ribosomes still scanning at this point may not initiate here, but pass to scan towards the more favourably situated AUG₁₄₈. The lack of detection of the second ORF polypeptide does not unequivocally mean that it is not expressed by the full-length P gene mRNA, it may be that the polypeptide is expressed in amounts that require long autoradiograph exposure for its detection.

8.3: THE GENERATION OF ANTISERUM AGAINST THE P GENE SECOND ORF PROTEIN PRODUCT.

8.3.1: The Qiagen Express system.

Antiserum specific against the PVM P gene second ORF protein product was generated by immunising rats with a synthetic oligopeptide representing amino acids 45 to 62 of the P2 amino acid sequence, as described in section 2.13.2. In order to test the anti-P2 protein antiserum for the ability to recognise second ORF polypeptide, the PVM P gene second ORF polypeptide was expressed in *E. coli* M15 [pREP4] using plasmid pQE32 of the Qiagen Express system (Qiagen Inc). This system can provide high level expression of protein which can be quickly and easily purified to a high level of homogeneity. cDNA inserts were cloned into these vectors such that their coding sequences were in-frame with an ATG initiation codon that initiated expression of a plasmid encoded peptide consisting of six histidine residues (6x histidine tag). The expressed fusion proteins therefore contain the 6x histidine tag linked to the amino-terminus. The plasmid constructs were transformed into *E. coli* K-12 strain M15 containing multiple copies of plasmid pREP4 which carries the *lac* I gene encoding the

lac repressor. The multiple copies of pREP4 present in the host cells ensure high levels of lac repressor and tight regulation of protein expression. Transcription in this system is controlled by a promoter/operator element consisting of the *E. coli* bacteriophage T5 promoter and two *lac* operator sequences. Expression is therefore inducible with IPTG. One-step purification, to greater than 95% homogeneity, is mediated by the extremely high affinity of the 6x histidine tag for Ni-NTA resin, a metal chelate absorbent.

8.3.2: Construction of plasmid pQE32.P2.

The PVM P gene second ORF coding region was amplified using PCR on PVM P gene cDNA. The oligonucleotide primers used for the amplification are shown below (both are in 5' to 3' polarity).

PVMP2A CCGGGATCCTCATGTCACCAAATATAACATGCCC

PVMP2B ACTAGATAGAATAGAGGAGAAGCTTAGG

Primers PVMP2A and PVMP2B hybridise to P gene nucleotides 128 to 153 and 552 to 572 respectively. The amplified fragment representing the P gene second ORF was isolated by agarose gel electrophoresis and purified by the fragment preparation technique. The fragment was end-filled using Klenow fragment, cleaved with restriction enzyme *Bam* HI and inserted into the *Bam* HI and *Sma* I sites present in the polylinker region of pQE32. These construction details are shown diagrammatically in figure 23. The resulting plasmid, pQE32.P2 was transformed into *E. coli* strain M15 [pREP4] and the PVM P gene second ORF polypeptide

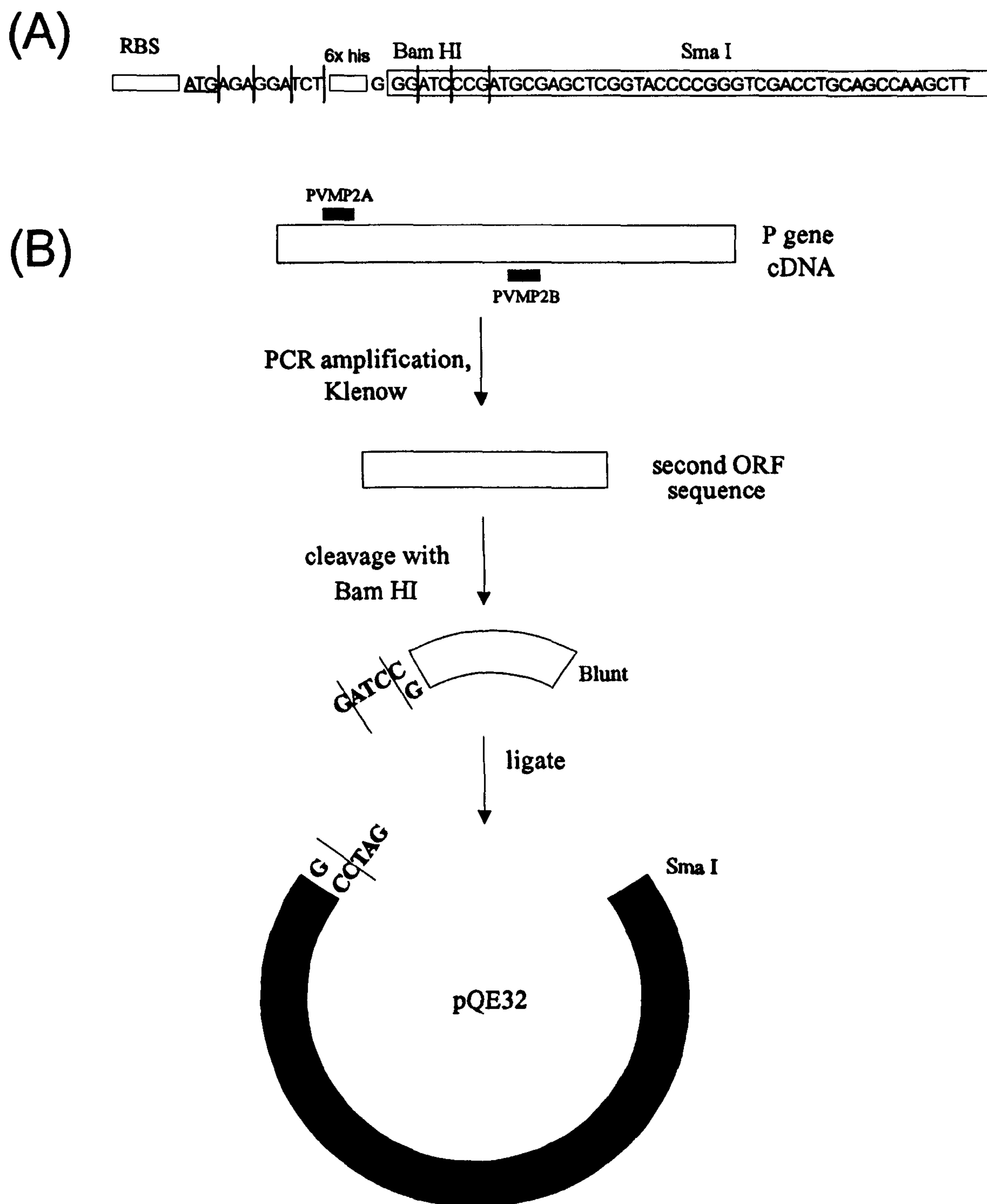


Figure 23 (A) and (B). Construction of pQE32.P2. (A) The multiple cloning site of pQE32 showing positions of the ribosome binding site (RBS), initiation codon (underlined) and the 6x his coding region. The coding triplets that are in-frame with the initiation codon are shown divided by vertical lines. (B). Construction details for plasmid pQE32.P2. PVM P gene second ORF coding region was amplified by PCR using oligonucleotide primers PVMP2A and PVMP2B (section 8.3.2). The amplified fragment was end-filled and cleaved using restriction enzyme *Bam* HI which cuts in PVMP1. The cleaved fragment was inserted into *Bam* HI and *Sma* I sites of pQE32. The coding triplets of the *Bam* HI recognition site are shown divided by vertical lines at the termini of the P2 fragment and the cleaved pQE32 vector.

product was expressed and purified as described in section 2.12.1

8.4: IMMUNOLOGICAL ACTIVITY OF ANTI-P2 ANTISERUM.

Antiserum taken from rats immunised with a synthetic peptide representing amino acids 45 to 62 of the PVM P gene second ORF protein product was tested for its ability to recognise second ORF protein product, expressed by *E. coli* M15 [pREP4] transformed with pQE32.P2, in Western blot analysis. The antiserum recognised a polypeptide of approximately 29 kDa, present only in the pQE32.P2 expression products. The pre-immune antiserum did not recognise this polypeptide (figure 24). In addition, the pre-immune antiserum did not recognise any other PVM specific polypeptides as determined by Western blot analysis against PVM-infected BSC1 cell material (and bacterially expressed PVM N protein). The molecular weight of the species recognised is greater than the observed molecular weight of the second ORF protein product expressed *in vitro* by translation in wheat germ lysate (20 kDa). However, the presence of the histidine leader affinity tag at the amino-terminus of the second ORF protein, encoded by the expression plasmid pQE32, increases the observed molecular weight of expressed species by approximately 6000 Daltons as described in section 9.2.2. Taking this into account, of the 29 kDa molecular weight observed for the *E. coli* M15 expressed second ORF fusion protein product, 6 kDa is likely to be due to the histidine tag and thus 23 kDa is likely to be due to the second ORF amino acid sequence. This calculated molecular weight for the bacterially expressed P2 protein is close to the observed molecular weight of P2 expressed in a wheat germ lysate (section 8.1).

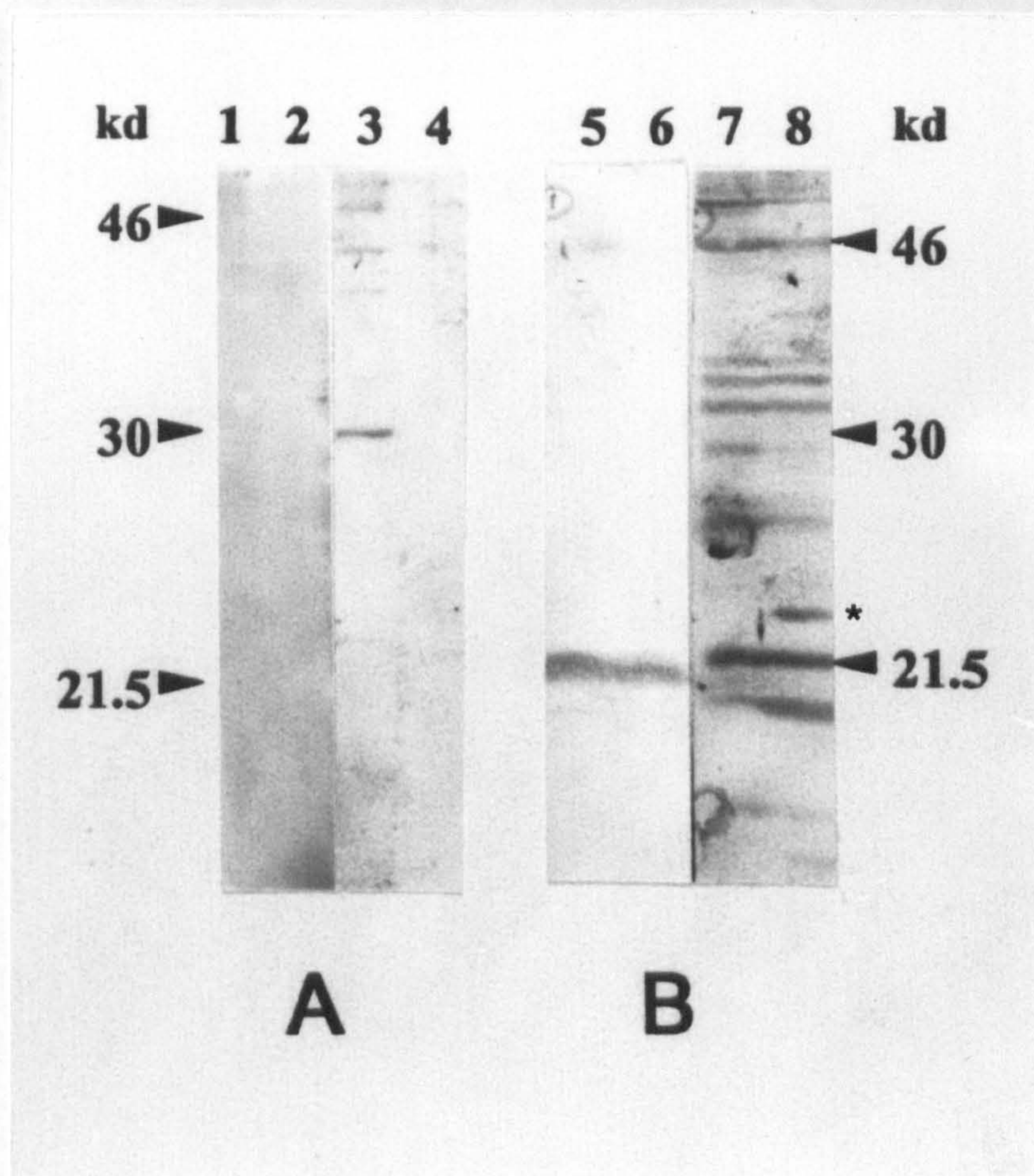


Figure 24. The PVM P gene second ORF encodes a polypeptide *in vivo*. Rabbits were immunised with a synthetic peptide representing amino acids 45-62 of the PVM P gene second ORF protein product. Success of the immunisation procedure was tested by the ability of post-immune serum to recognise purified second ORF protein product in Western blot analysis. Second ORF product was expressed using *E. coli* M15 [pREP4] which was transformed with the IPTG-inducible expression plasmid pQE32.P2 which contained the P gene second ORF coding sequence (see section 8.3.2 for construction details). Plasmid pQE32 did not contain the second ORF coding sequence and was used as a control. Lane 1: Products of pQE32.P2 expression blotted with pre-immune sera. Lane 2: Product of pQE32 expression blotted with pre-immune sera. Lane 3: Expression products of pQE32.P2 blotted with post-immune sera. Lane 4: Expression products of pQE32 blotted with post-immune sera. The positions of molecular weight markers are indicated at the sides of the blot. In this gel system, the bacterially expressed second ORF protein product migrates with an apparent molecular weight of 29 kDa due to the presence of an amino-terminal affinity tag encoded for by the pQE32.P2 plasmid which facilitates purification of the species. The pre- and post-immune sera were then used in Western blot analysis with PVM-infected and mock-infected BSC1 cell material. Lane 5: Mock-infected cell material blotted with pre-immune sera. Lane 6: PVM-infected cell material blotted with pre-immune sera. Lane 7: Mock-infected BSC1 cells blotted with post-immune sera. Lane 8: PVM-infected cells blotted with post-immune sera. The Western blot analysis identified a 23 kDa polypeptide that was only present in the PVM-infected cell material, and was identified only with post-immune serum (*)

8.5: IDENTIFICATION OF P2 IN PVM-INFECTED BSC1 CELLS.

Figure 24 shows a Western blot on which the expression products of PVM-infected and mock-infected BSC1 cell extracts were blotted against sera taken from rats both before and after immunisation with the synthetic peptide. A polypeptide with a molecular weight of approximately 23 kDa Daltons was identified in the PVM-infected BSC1 cell material by the post-immune serum but not the pre-immune serum. This 23 kDa species was not present in the mock-infected BSC1 cell material. The molecular weight of this species is similar, but not identical, to the observed molecular weight of the P gene second ORF protein product translated *in vitro* using wheat germ lysate which is approximately 20 kDa. However the molecular weight of this PVM-infected BSC1 cell species is the same as the calculated molecular weight of P2 expressed from pQE32.P2 which is approximately 23 kDa (section 8.4).

8.6: DISCUSSION

The discrepancy between the observed molecular weight of wheat germ expressed second ORF product (20 kDa) and the PVM-infected BSC1 cell polypeptide that is thought to represent the second ORF product (23 kDa) is cause for this assignment to be made tentatively. The recognition by the antiserum of both bacterially expressed P2 and the PVM-infected BSC1 cell expressed P2 polypeptide dictates that these species share polypeptide sequences. It is possible that the calculated difference in the apparent molecular weights of these species is due to a modification that occurs to the *in vivo* expressed second ORF polypeptide in the host BSC1 cells. However, the possibility that the 23 kDa species is expressed by either PVM or by the host BSC1 cells in response to the infection, and cross-reacts with the post-immune serum due to

the presence of a common peptide motif must not be overlooked, although this is unlikely.

The presence of an accessible coding region of considerable size, as is represented by the PVM P gene second ORF, would suggest that this is indeed likely to be used for the expression of a polypeptide because production of an unnecessary polypeptide would presumably burden the virus and so the presence of the useless coding potential would be selected against. This argument would support the idea that *in vivo*, the P gene second ORF expresses a polypeptide species, and that it is functional. However, the first initiation codon on the second ORF is in poor nucleotide context (section 8.2). It may be that the context is sufficiently poor to prevent significant expression of the second ORF product and if this species had no function, then the second ORF coding region could exist without burdening the replicative efficiency of the virus.

The 23 kDa polypeptide identified by Western blotting of PVM-infected BSC1 cells using the anti-second ORF antiserum does not correspond to the 23 kDa polypeptide expressed from the P protein ORF of the P gene, which was identified by antibody 26/3/B5 in Western blot analysis. If both were the same then the anti-peptide antiserum would also recognise the more abundant full-length P protein since the amino acid coding sequence of the 23 kDa P protein related polypeptide is also present in the full-length P protein.

As was mentioned in section 7.1, the monoclonal antibody 26/3/B5 identifies a total of eleven PVM-infected BSC1 cell protein products and of these eleven, seven are still uncharacterised. Several possibilities regarding the identity of these species are presented in section 7.3.2. There is another possibility and this is that these uncharacterised P protein related polypeptides may be composed, in part, of amino acid sequence that is encoded for by both the P protein ORF and also the second ORF as a result of polymerase slippage during transcription. The only PVM-infected BSC1

cell polypeptide recognised by the anti-peptide antiserum was the 23 kDa species, and this observation suggests that no other PVM-infected BSC1 cell polypeptides have second ORF amino acid sequences (from the Western blot shown in figure 24 the second ORF antiserum may appear to recognise PVM infected cell polypeptides with molecular weights of approximately 19 kDa and 40 kDa, however this is due to poor reproduction of the original blot which has failed to show a comigrating band in the mock infected cell material lane [lane 7]). However, since the immunogen used to raise the second ORF serum was a short peptide, then PVM-infected BSC1 cell polypeptides representing second ORF amino acid sequences outside this small ORF would not be identified with this serum. There are two ways in which a polypeptide encoded by parts of both ORFs can be expressed and be recognised by monoclonal antibody 26/3/B5 but not the anti-second ORF polypeptide antiserum. Firstly, the polypeptide would have to be initiated from AUG₁₁₁ and the frame shifting occur at a site upstream of that used to encode the region of P2 represented by the synthetic peptide. Alternatively, the polypeptide would have to be initiated from an AUG downstream of the region which encodes the sequence of the synthetic peptide and the frame shifting occur at some point within the remainder of the P2 coding region. These theoretical modes of polypeptide expression would be unlike that seen in any other *Paramyxoviridae* family member and are very unlikely to occur. As mentioned in section 3.4.6, the PVM P gene does not contain a sequence that is known to promote polymerase slippage and thus if polypeptides are expressed by PVM which contain amino acid sequence from the two ORFs then the mechanism by which the ORFs are switched is probably not the same as for the *Paramyxoviridae* family members. A more informative picture of the expression of the PVM P gene may be obtained by generating second ORF antiserum by immunising animals with a bacterially expressed version of the complete second ORF coding region. This would mean that any PVM-infected BSC1 cell polypeptide that possessed second ORF sequence could be identified.

CHAPTER 9

THE POLYPEPTIDE EXPRESSION PRODUCTS OF THE PVM N GENE.

9.1: PVM N GENE EXPRESSION *IN VITRO*.

9.1.1: Construction of transcription plasmids pGEMN and pGEMNTR.

A cDNA fragment containing the entire coding sequence of the PVM N gene was excised from plasmid 6.19 (Barr *et al.*, 1991) by digestion with restriction enzymes *Fok* I and *Bam* HI. This cDNA segment was isolated by agarose gel electrophoresis and purified using the fragment preparation technique. The fragment was blunt ended with Klenow fragment and inserted into the *Sma* I site of the T7 transcription vector pGEM1. The resulting plasmid was named pGEMN. A transcription vector, pGEMNTR, was also constructed which contained nucleotides 592 to 1215 of the PVM N gene. This was achieved by excising and blunting the full-length N gene cDNA from clone 6.19 as described above, digesting the fragment with restriction enzyme *Dra* I and then inserting it into the *Sma* I site of pGEM1. The details involved in the construction of these plasmids is shown in figure 25.

9.1.2: *In vitro* translation products of pGEMN and pGEMNTR.

Messenger RNA transcribed from pGEMN was translated in a rabbit reticulocyte lysate system in the presence of ³⁵S methionine and the radiolabelled expression products were analysed by SDS-PAGE and autoradiography (figure 26). The N gene coding region of pGEMN directed the *in vitro* expression of two major polypeptides, the 43 kDa full-length N protein and a 24 kDa polypeptide. In order to establish the origin of the 24 kDa polypeptide, the transcription vector pGEMNTR was constructed (section 9.1.1) which contains nucleotides 592 to 1215

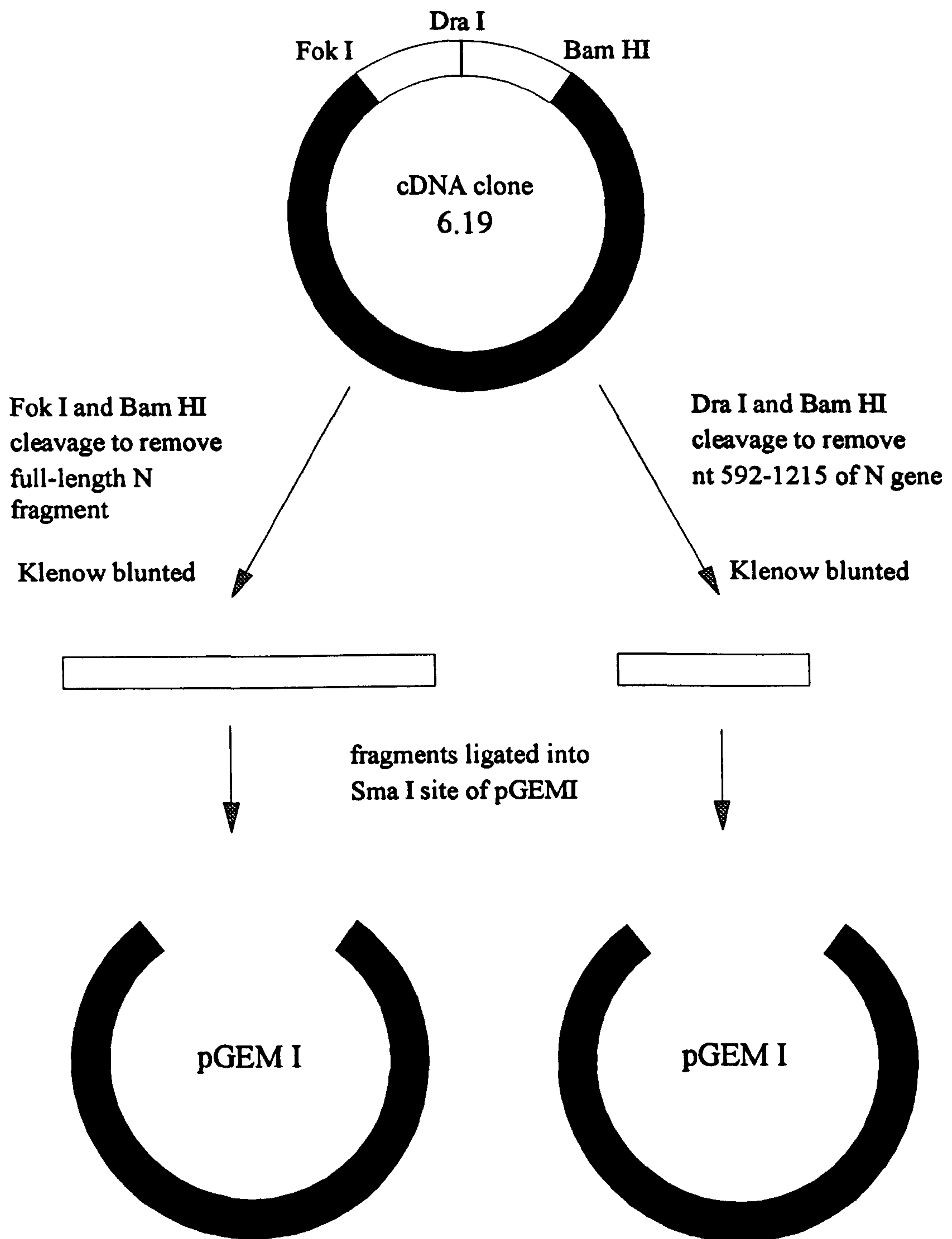


Figure 25. Construction of pGEMN and pGEMNTR. (a). A fragment representing full-length N gene cDNA was excised from cDNA clone 6.19 (Barr *et al.*, 1991) by digestion with restriction enzymes *Fok* I and *Bam* HI. The fragment was blunt ended using Klenow and inserted into the *Sma* I site of transcription plasmid pGEMI to derive the plasmid pGEMN. **(b).** A fragment representing nucleotides 592-1215 of the PVM N gene cDNA was removed from clone 6.19 using restriction enzymes *Dra* I and *Bam* HI. The fragment was blunted using Klenow and inserted into the *Sma* I site of transcription plasmid pGEMI to derive pGEMNTR.

of the N gene cDNA coding region. mRNA transcribed from pGEMNTR was translated in a rabbit reticulocyte system and analysis by SDS-PAGE revealed that pGEMNTR directs the *in vitro* synthesis of the 24 kDa N protein species alone (figure 26).

The first AUG codon on the mRNA transcribed from this plasmid is located at a position corresponding to nucleotides 612 to 614 of the full-length N gene and this AUG codon is not capable of directing the synthesis of a polypeptide due to the presence of multiple downstream stop codons in this reading frame (Barr *et al.*, 1991). The next AUG codon downstream is located at nucleotides 629 to 631 of the N gene and this AUG codon is in the same open reading frame as the large ORF that encodes the PVM N protein. Initiation of translation at AUG₆₂₉ is predicted to result in the translation of a polypeptide with a molecular weight of 21,478 Da. There are no other AUG codons sufficiently near to AUG₆₂₉ that could initiate expression of a polypeptide of 24 kDa. The next in-frame AUG codon downstream of AUG₆₂₉ is at nucleotide positions 720 to 722 which is predicted to initiate synthesis of a 16,204 Da polypeptide. There are no other alternative initiation codons, such as the ACG codon used to initiate translation of the Sendai virus C' polypeptide (Gupta and Patwardhan, 1988; Curran and Kolakofsky, 1988a), that could give rise to the 24 kDa polypeptide, and therefore AUG₆₂₉ must be responsible for initiation of the 24 kDa species. Because of the abundance of the 24 kDa species *in vitro*, and its potential as a possible regulatory protein, a study was undertaken to see if the 24 kDa species could be immunologically detected in PVM-infected BSC1 cell material.

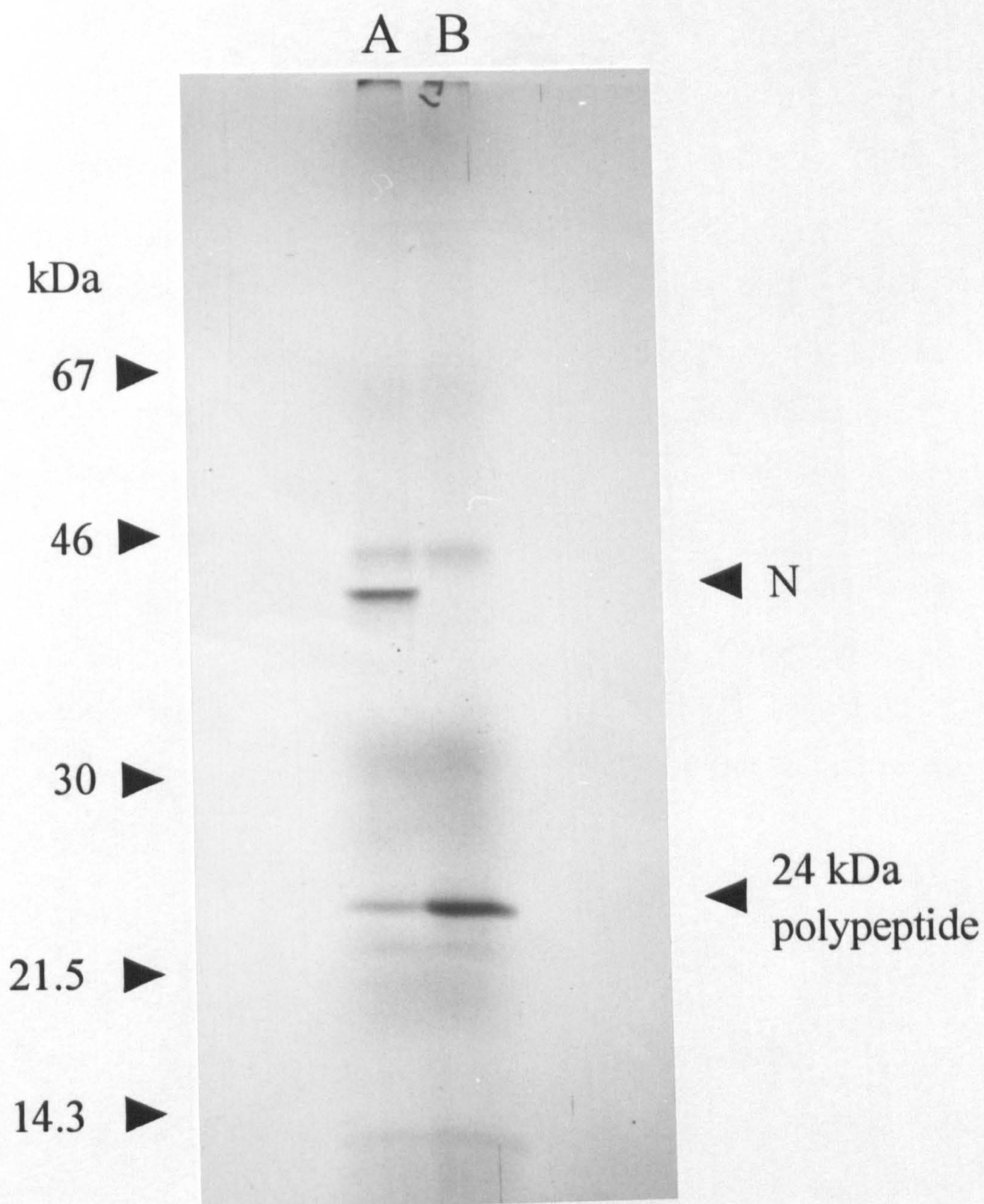


Figure 26. The full-length N gene cDNA directs the *in vitro* expression of the full-length N protein (43 kDa) and a amino-terminally truncated polypeptide with a molecular weight of 24 kDa. mRNA transcribed from pGEMN (full-length N gene cDNA insert) and pGEMNTR (cDNA insert with nucleotides 592-1215 of the N gene) was used to programme a reticulocyte lysate system and the radioactively labelled expression products were subjected to SDS-PAGE and autoradiography. Lane A: mRNA transcribed from pGEMN. Lane B: mRNA transcribed from pGEMNTR. The positions of molecular weight markers are indicated on the left.

9.2: GENERATION OF ANTISERUM AGAINST THE 24 kDa PVM N PROTEIN SPECIES.**9.2.1: Construction of pQE32.24.**

A cDNA fragment possessing nucleotides 621-1215 of the PVM N gene was generated by cleaving the full-length N gene cDNA fragment with the restriction enzymes *Kpn* I and *Hinc* II. The DNA segment was isolated by agarose gel electrophoresis and then purified using the fragment preparation procedure. The fragment was inserted into the *Kpn* I and *Sma* I sites of expression vector pQE32 of the Qiagen Express system (Qiagen Inc, described in section 2.12.1) such that the N gene open reading frame was in-phase with the pQE32 ATG initiation codon specifically intended for the expression of inserted DNA sequences. The construction of this plasmid is shown in figure 27. The resulting vector was designated pQE32.24. The plasmid was transformed into *E. coli* strain M15 [pREP4] (Qiagen Inc.) by the transformation procedure described in section 2.5.2.

9.2.2: Immunisation of rabbit with purified 24 kDa N protein species.

The expression product of pQE32.24 was purified according to the procedure described in section 2.12.1, and subjected to SDS-PAGE to confirm the success of the purification procedure (figure 28). The purified polypeptide was used to generate antiserum in rabbits as described in section 2.13.1. Figure 29 shows a Western blot on which the expression products of IPTG induced and un-induced cultures of *E. coli* M15 [pREP4] transformed with pQE32.24 were reacted with the sera taken from rabbits both before and after immunisation with the purified 24 kDa N protein species. The post-immune antiserum reacted strongly with a polypeptide with an

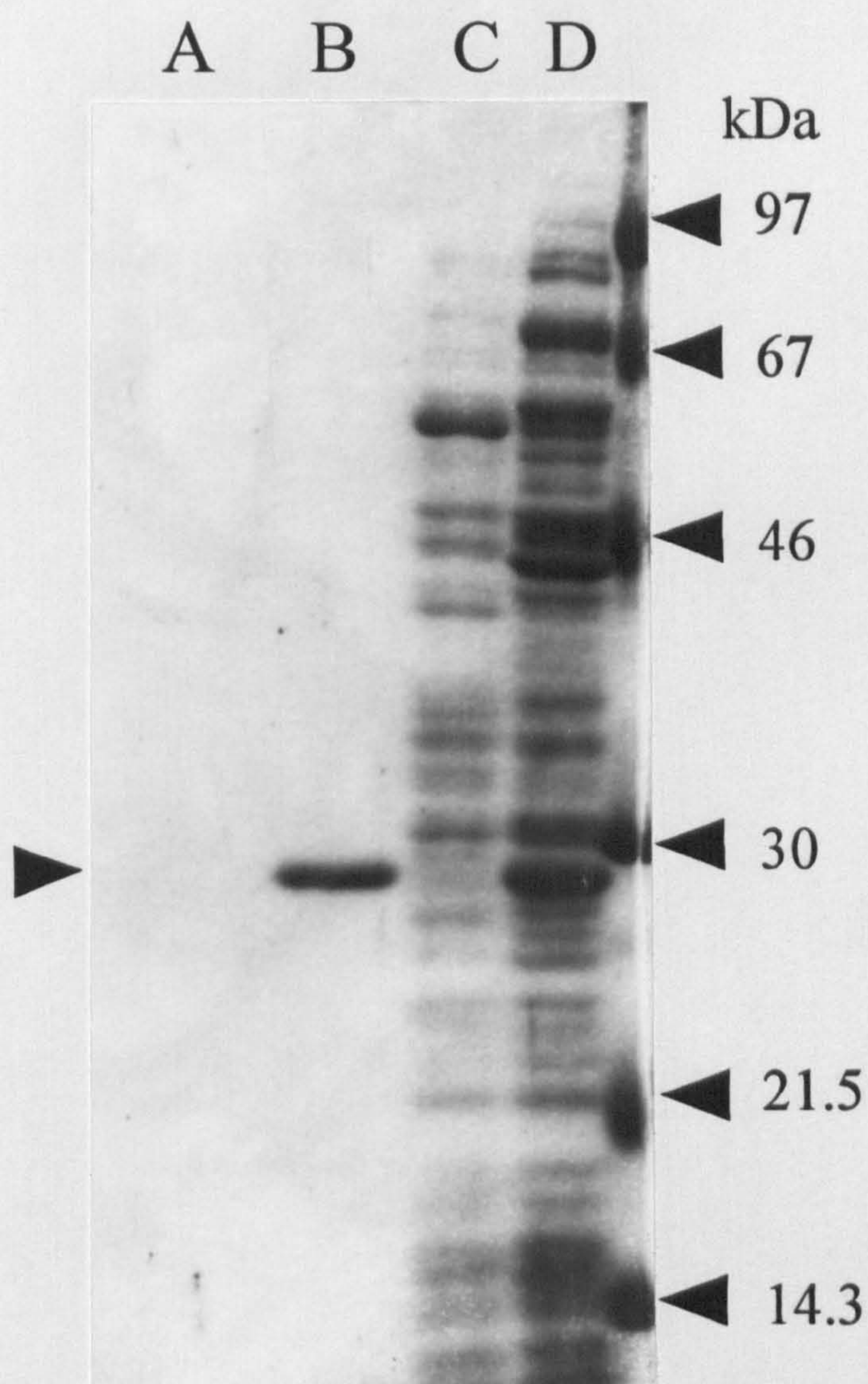


Figure 28. Expression and purification of the truncated N protein using the Qiagen Express system. The products of bacterial expression were subjected to SDS-PAGE analysis before and after purification using Ni-NTA resin. Lane C: Total expression products of un-induced *E. coli* M15 [pREP4] transformed with pQE32.24. Lane D: Total expression products of IPTG-induced *E. coli* M15 [pREP4] transformed with pQE32.24. Lane A: Expression products of un-induced *E. coli* M15 [pREP4] transformed with pQE32.24 after purification using Ni-NTA resin. Lane B: Expression products of IPTG-induced *E. coli* M15 [pREP4] transformed with pQE32.24 after purification using Ni-NTA resin. The position of molecular weight markers are shown on the right, and the arrow on the left indicates the position of the purified truncated N protein

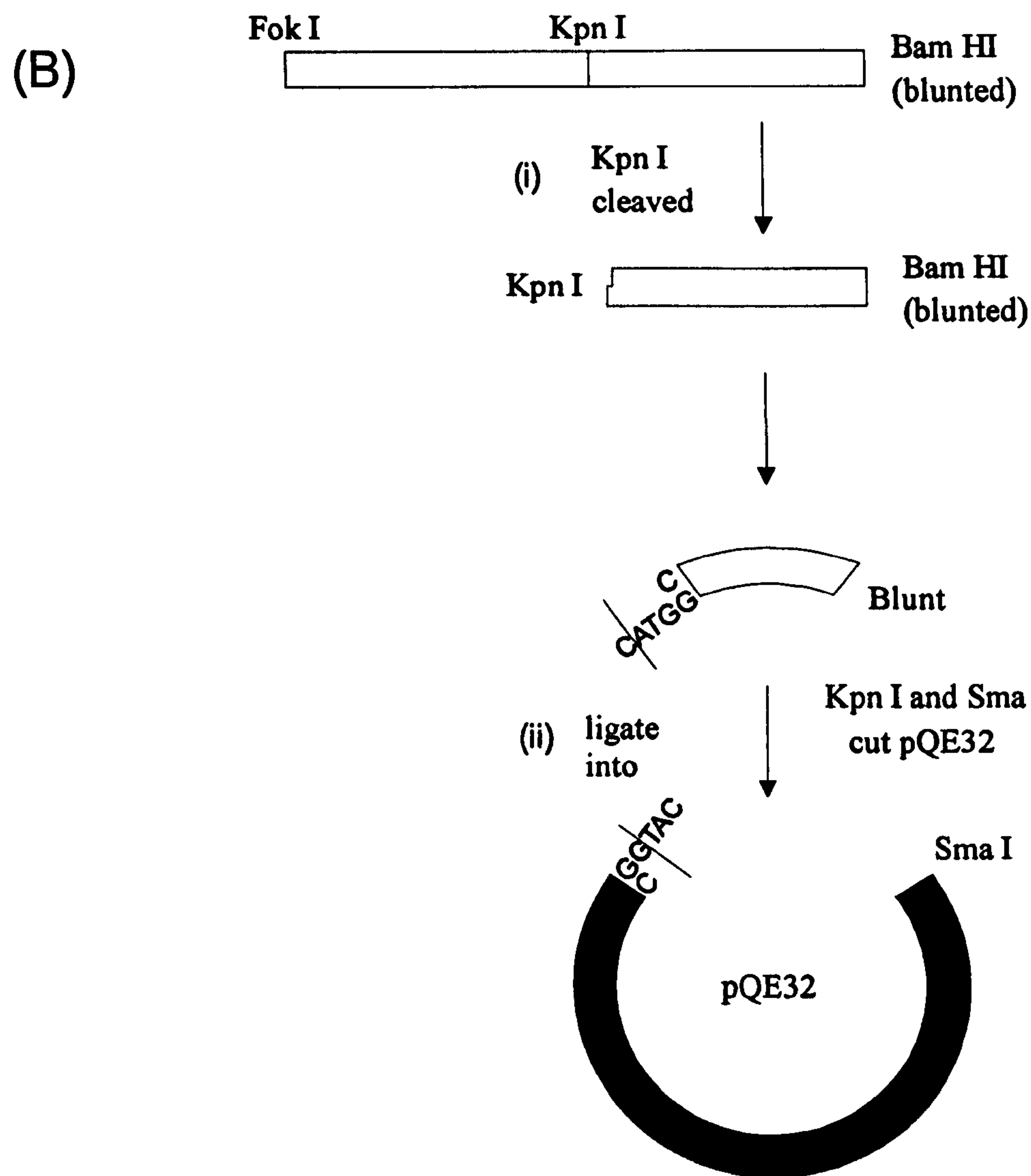


Figure 27 (A) and (B). Construction of plasmid pQE32.24. (A) The multiple cloning site of pQE32 showing positions of the ribosome binding site (RBS), the initiation codon (underlined), and the 6x his coding region. The coding triplets that are in-frame with the initiation codon are shown divided by vertical lines. (B). Construction details for pQE32.24. (i). A full-length N gene cDNA fragment was cleaved with restriction enzyme *Kpn* I. (ii). The *Kpn* I-*Bam* HI (blunt) fragment which represented nucleotides 621-1215 of the PVM N gene, was inserted into the *Kpn* I and *Bam* HI (blunted) sites of pQE32.

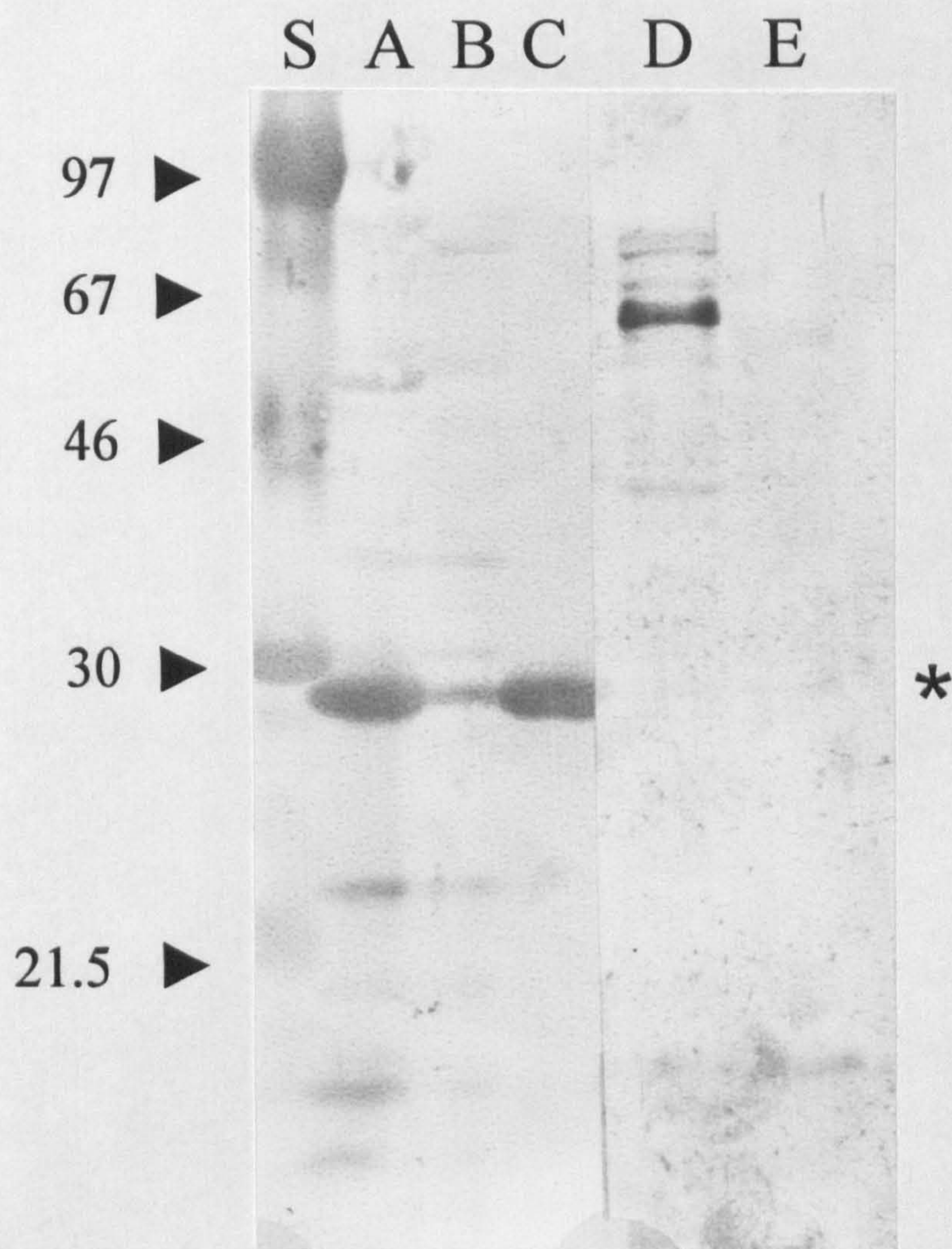


Figure 29. Antiserum was raised against purified 24 kDa N protein species expressed in *E. coli* M15 [pREP4] transformed with pQE32.24. Post-immune serum was used in Western blot analysis against the expression products of *E. coli* M15 [pREP4] transformed with pQE32.24. Lane A: IPTG-induced total bacterial culture. Lane B: Un-induced total culture. Lane C: IPTG-induced culture after purification with Ni-NTA resin. Pre-immune rabbit serum used in Western blot analysis with expression products of *E. coli* M15 [pREP4] transformed with pQE32.24. Lane D: IPTG-induced total bacterial culture. Lane E: IPTG-induced bacterial culture after purification with Ni-NTA resin. The position of the bacterially expressed 24 kDa truncated N protein is marked (*). The position of molecular weight markers (lane S) is indicated on the left.

observed molecular weight of 30 kDa in the induced pQE32.24 extract. The PVM N gene coding sequence contained within pQE32.24 is identical to that contained in pGEMNTR which directs the *in vitro* expression of the 24 kDa N protein species (section 9.1.2). The observed molecular weight difference between the 30 kDa expression product of pQE32.24 and the 24 kDa *in vitro* expression product of pGEMNTR is therefore due to the presence of the 6x histidine tag. The antiserum did not identify a corresponding polypeptide in the induced pQE32 extract. The pre-immune serum did not recognise the purified 24 kDa species, similarly, the antiserum did not react with a protein in the expression products of un-induced *E. coli* M15 [pREP4] containing pQE32.24.

9.2.3: Identification of the 24 kDa species in PVM-infected BSC1 cell material.

The anti-24 kDa N protein antiserum was used in Western blot analysis with PVM-infected and mock-infected BSC1 material (figure 30). In the PVM-infected BSC1 cell material, the antiserum identified a 43 kDa species representing the PVM N protein and also a polypeptide with an observed molecular weight of approximately 30 kDa. Neither of these two species were identified in the mock-infected BSC1 cell material. A polypeptide with a molecular weight of 24 kDa was not identified in the PVM-infected BSC1 cell material. The identity of the 30 kDa species is unknown although due to its recognition by the antiserum, it must represent a region of the PVM N protein that includes a portion of its carboxyl-terminal half.

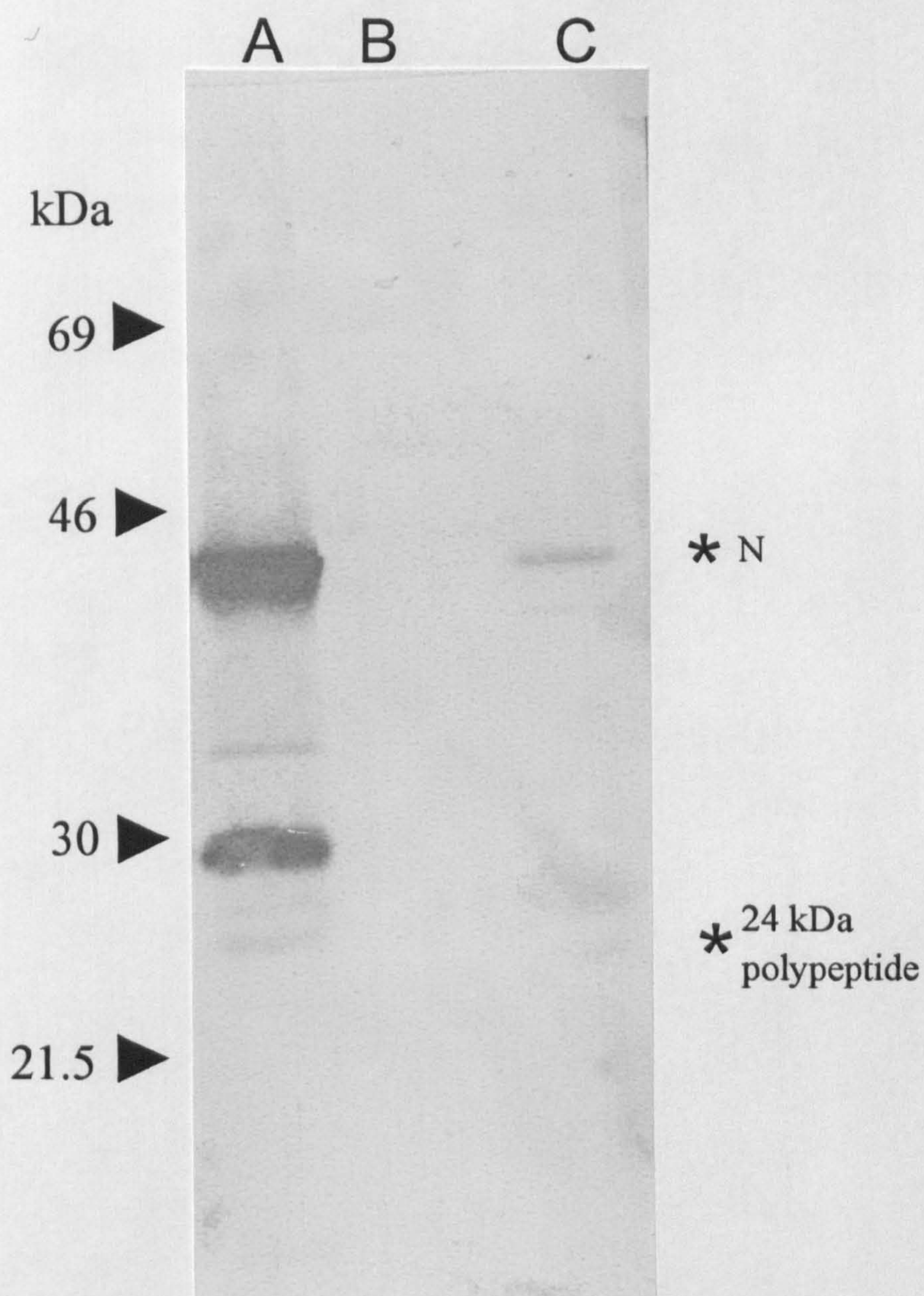


Figure 30. The N protein and a N protein related 30 kDa polypeptide are present in PVM-infected BSC1 cells. The anti-24kDa polypeptide antiserum was used in western blot analysis with PVM-infected BSC1 cell material (lane A) and mock-infected BSC1 cell material (lane B). The PVM N protein (43 kDa) and also a polypeptide with a molecular weight of 30 kDa were identified only in the PVM-infected BSC1 cells. For comparison, the 43 kDa full-length and 24 kDa truncated N proteins expressed *in vitro* using reticulocyte lysate programmed with pGEMN were also electrophoresed (lane C), and the positions of these polypeptides are indicated (*) on the left. The positions of molecular weight markers are indicated on the left.

9.3: DISCUSSION

9.3.1: Expression products of the PVM N gene.

Rabbit reticulocyte lysate programmed with mRNA transcribed from pGEMN synthesises the full-length N protein and also a 24 kDa species which, by expression of mRNA transcribed from the truncated N gene cDNA of plasmid pGEMNTR, was shown to be initiated from the AUG codon at nucleotide position 629 to 631.

9.3.2: Expression of the 24 kDa N protein species.

There are several explanations that can account for the expression of the 24 kDa polypeptide. Firstly it may be made as a consequence of the proposed leaky scanning mechanism (Kozak, 1986a; 1986b; 1986c; 1989) by 40S ribosome subunits that fail to initiate at the preceding initiation codons. Secondly, although extremely unlikely, it may be expressed as a result of mRNA breakage that occurs in such a place as to make AUG₆₂₉ the 5' proximal AUG codon on the broken mRNA and thus virtually ensure its usage (Kozak, 1987). Finally, it may also be generated as a result of proteolytic cleavage of the full-length N protein.

If leaky scanning is responsible for initiation of translation on AUG₆₂₉, the 40S ribosomal subunit present in the reticulocyte lysate must pass and fail to initiate on the preceding sixteen AUG codons of the N gene cDNA many of which are in a different reading frame and are incapable of directing the synthesis of large polypeptides. The mechanism of leaky scanning appears to be heavily dependent on the nucleotide sequence surrounding the codons that could potentially initiate translation. The nucleotide sequence surrounding the first initiation codon on the PVM N mRNA is such that the context is sub-optimal (Kozak, 1986a; 1986b; Figure 31). Consequently, in accordance with the leaky scanning hypothesis, a proportion of

AUG codon number	reading frame	sequence surrounding AUG codon			
1	2	<u>CC</u> AAA	AUG	U	*
2	3	G <u>C</u> U <u>C</u> A	AUG	A	
3	3	<u>C</u> AA <u>A</u> UG	AUG	U	
4	3	AG <u>G</u> <u>C</u> G	AUG	U	
5	3	UG <u>G</u> U <u>C</u>	AUG	C	
6	2	AU <u>G</u> <u>C</u> U	AUG	C	*
7	2	U <u>C</u> <u>G</u> <u>G</u> <u>C</u>	AUG	U	*
8	3	<u>CC</u> AAU	AUG	C	
9	2	AU <u>G</u> <u>CC</u>	AUG	U	*
10	3	<u>C</u> AGAG	AUG	A	
11	3	<u>C</u> U <u>A</u> CA	AUG	U	
12	1	GU <u>A</u> AA	AUG	U	
13	3	AAUAG	AUG	<u>G</u>	
14	3	GAG <u>G</u> <u>C</u>	AUG	A	
15	3	G <u>C</u> AUG	AUG	C	
16	3	A <u>C</u> UUG	AUG	C	
17	3	UUUAA	AUG	U	
18	2	<u>CC</u> A <u>A</u> <u>C</u>	AUG	<u>G</u>	*

Figure 31. The nucleotide context that surrounds the first 18 initiation codons of the PVM N gene mRNA. For each of the N gene AUGs listed above, the nucleotide identities with the optimal context motif determined to be CCA/GCCAUGG (Kozak, 1986a) are shown in bold face and are underlined. The initiation codons that are in the N protein ORF (ORF 2) are identified with an asterisk (*).

40S ribosomal subunits will initiate at the first AUG of the N gene mRNA, but since it is sub-optimally situated, some will pass along the mRNA to the next AUG. The context of the next sixteen AUG codons are also poor (figure 31), thus while some ribosomes may initiate at these codons, some will pass by and scan downstream towards the next. For initiation to occur on the eighteenth initiation codon, AUG₆₂₉, the proportion of ribosomes that pass all the preceding AUG codons must be vastly greater than the proportion that would stop and initiate, otherwise no ribosomes would be left scanning on the mRNA to initiate at AUG₆₂₉. The highly favourable sequence surrounding AUG₆₂₉ is such that the 40S ribosomal subunits that have reached this far will probably proceed no further. This suggestion is supported by the observation that no low molecular weight polypeptides are expressed *in vitro* from mRNA transcribed from pGEMN. Secondary structure prediction analysis for the PVM N gene mRNA indicates that AUG₆₂₉ is located on an exposed and isolated loop structure, whereas the previous three AUGs of the N protein ORF are positioned within stem formations (figure 32). However it is thought that presence of buried 5' AUG triplets cannot direct a 40S ribosome subunit to initiate at downstream AUGs (Kozak, 1986c). The proposed *in vitro* expression of the 24 kDa polypeptide from the eighteenth AUG codon is therefore probably due to the consequences of local sequences surrounding the internal AUG triplets alone. Whether the proposed phenomenon of reading frame commitment (Kozak, 1989 and references therein) could play a part in the expression of the 24 kDa polypeptide is not known. This phenomenon suggests that the ribosomes only scan in one of the three ORFs because they move along the mRNA in steps of three nucleotides. If the phenomenon of reading frame commitment did operate in the translation of this message, then the number of AUGs that the ribosome had to scan past would be reduced from sixteen to four, thus making initiation at AUG₆₂₉ considerably more likely. The abundance of the 24 kDa species in the expression products of reticulocyte lysate programmed with pGEMN transcribed mRNA suggests that initiation of translation on AUG₆₂₉ is a

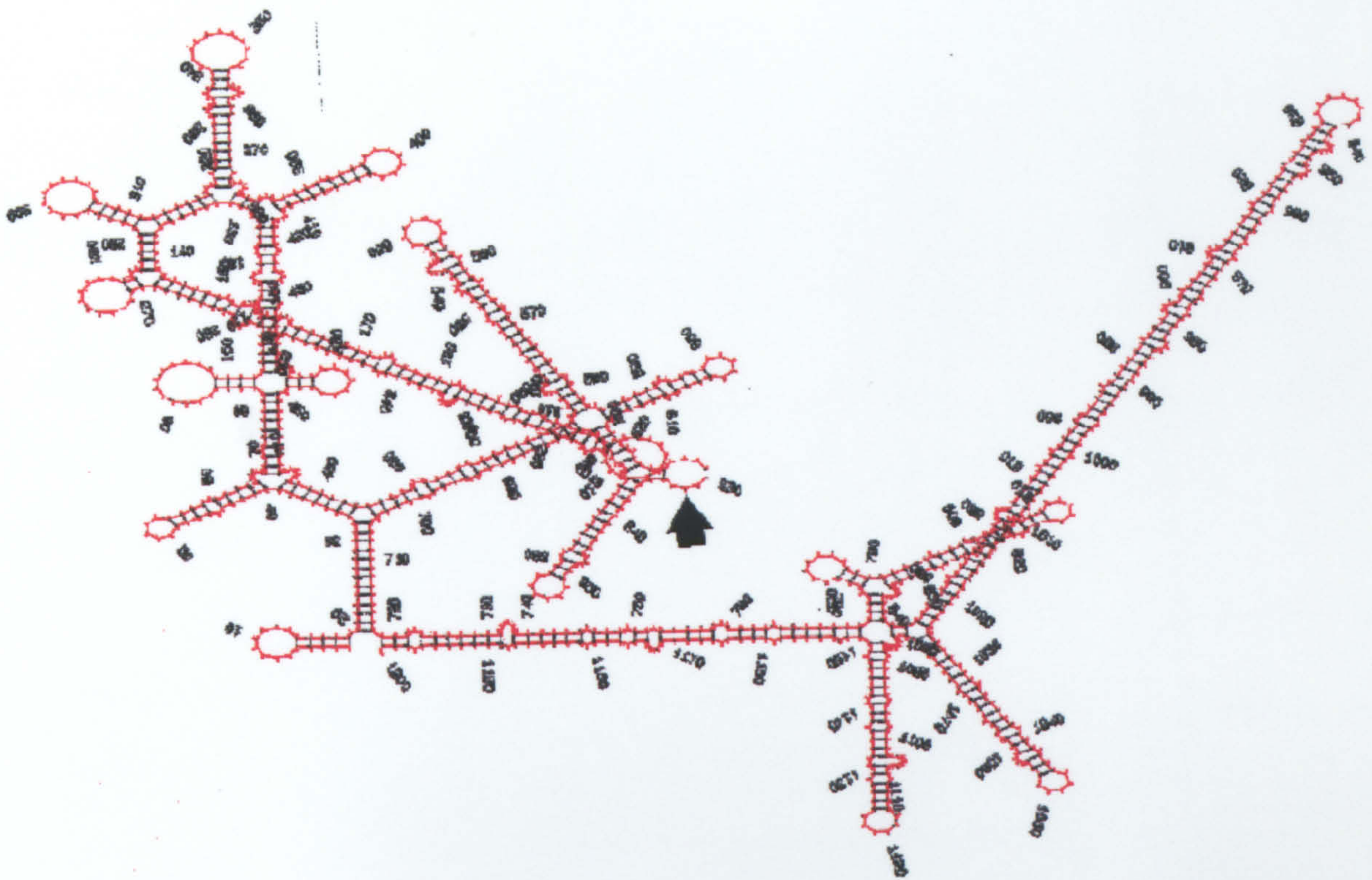


Figure 32. SQUIGGLES plot of the PVM N gene mRNA using the FoldRNA programme (Zuker, 1989). The AUG codon responsible for the initiation of the 24 kDa polypeptide *in vitro* (AUG₆₂₉) is positioned on an exposed loop structure (indicated with the arrow) whereas the previous four AUG codons of the N protein ORF (at nucleotide positions 32, 149, 179 and 242) are buried within stem formations. Whether the location of AUG₆₂₉ influences the ability of scanning ribosomes to initiate at this position is unknown, although it has been suggested that the presence of buried 5' AUG triplets cannot direct a ribosome to initiate at downstream AUGs (Kozak, 1986c).

common event and so whatever method of scanning is used it appears that a larger number of ribosomes do initiate at AUG₆₂₉. Also interesting is the apparent lack of detection of any N protein related species that could be expressed from the previous three upstream in-frame AUG codons located at positions 149, 179 and 242, either in western blot analysis using the polyclonal anti-24 kDa N protein antiserum with PVM-infected BSC1 cell material, or with *in vitro* translations using N gene mRNA. These three seemingly inactive codons are all situated in a nucleotide context that is considered to be poor for translational initiation since whilst they have either an A or a G at position -3, they do not have a G at the critical +4 position (Figure 31). The context of these codons is similar to that of the first AUG codon of the major ORF, but this is presumably used due to its position at the 5' end of the mRNA fragment. It is possible that the high frequency of initiation at AUG₆₂₉ is a consequence of these preceding codons being unable to support translational initiation at a high frequency.

Another possible, but highly unlikely, reason for expression of the 24 kDa polypeptide is mRNA breakage. If mRNA breakage was responsible for causing the expression of the 24 kDa polypeptide then the breakages would have to position AUG₆₂₉ at the 5' end of an mRNA fragment and thus the cleavage event would have to be at a position somewhere between nucleotides 400 and 628. This would mean that there would be, in addition to the mRNA fragment that had AUG₆₂₉ at its 5' end, another mRNA with AUG₃₂ at its 5' end, which would be capable of directing the translation of polypeptides with approximate molecular weights of between 13 kDa and 20 kDa depending on the position of cleavage. No such polypeptides are ever seen in the SDS-PAGE analysed *in vitro* expression products of pGEMN mRNA. If the cleavage event occurred such that either of the second, third or fourth N gene long ORF polypeptides were at the 5' end of a cleaved mRNA fragment then expression of AUG₆₂₉ would be more favourable but at the same time would lead to the expression of a polypeptide from one of these upstream AUGs, and, as was mentioned earlier,

polypeptides expressed from these AUGs are never seen. If a proteolytic cleavage event was responsible for generation of the 24 kDa polypeptide *in vitro*, then the cleavage would have to occur at the precise position on the PVM N protein that corresponded to AUG₆₂₉ and this is extremely unlikely. Additionally, if cleavage occurred at this position, then a 17 kDa species representing the amino-terminus of N protein would also be generated, and a species of this size has not been identified in *in vitro* translations.

9.3.3: The 30 kDa N protein species expressed in PVM-infected BSC1 cells.

Antiserum was raised against bacterially expressed PVM 24 kDa N protein related polypeptide (section 2.13.1). In PVM-infected BSC1 cell material, two polypeptide species encoded by the N gene were detected by the anti-24 kDa antiserum in Western blot analysis, the full-length 43 kDa N protein, and a 30 kDa species. A polypeptide with a molecular weight of 24 kDa was not recognised by this antibody in either PVM-infected or mock-infected BSC1 cells (figure 30). There are several possibilities regarding the identity of the 30 kDa species. Firstly, it may be a breakdown product of the PVM N protein generated by a specific cleavage event. The cleavage event would have to be highly specific and only occur at one position to account for the total absence of any other polypeptides of a molecular weight smaller than the full-length N protein species. A 30 kDa species is not seen following *in vitro* translation of the PVM N protein, and so the putative cleavage event would probably be mediated by cellular agents.

Alternatively, the 30 kDa species may be the product of an initiation event on an internal N gene AUG codon. There is no initiation codon present on the PVM N gene capable of initiating the synthesis of a polypeptide having a predicted molecular weight of 30 kDa. The two AUG codons that are capable of encoding polypeptides

closest to the 30 kDa species are AUG₆₂₉, which *in vitro* encodes the 24 kDa species, and the next AUG codon upstream of this, located at nucleotide positions 242 to 244, which could potentially encode a polypeptide with a predicted molecular weight of approximately 36 kDa. Clearly, if internal initiation from either of these AUGs is responsible for the generation of the *in vivo* 30 kDa species the translation product must be modified in some way that changes its observed molecular weight. The 30 kDa species is not seen following *in vitro* translation of the PVM N protein, and thus if modification is responsible for the presence of this polypeptide, then the modification would appear to be a process mediated by cellular factors. As was mentioned above, *in vitro* studies indicate that the 24 kDa polypeptide is made in high abundance which suggests that initiation on AUG₆₂₉ is a common event. In contrast to this, a 36 kDa polypeptide product is never seen when mRNA transcribed from pGEMN is expressed *in vitro* suggesting that initiation of translation on AUG₂₄₂ seems to be a rare occurrence. On the basis of these *in vitro* observations, if initiation on an internal AUG codon was responsible for the generation of the 30 kDa species AUG₆₂₉ is probably the most likely candidate. It is well known that *in vitro* translation using the rabbit reticulocyte system can lead to the synthesis of artefactual polypeptides due to aberrant initiation events (Kozak, 1989 and references therein). This is due in part to the willingness of reticulocyte lysate 40S ribosomal subunits to use uncapped mRNAs as their substrate. It is possible therefore that the *in vitro* results do not give an accurate reflection of the abilities of the PVM N gene AUG codons to support translational initiation *in vivo*.

Western blot analysis demonstrated within the estimation limits permissible using this technique, that the 30 kDa species is present in PVM-infected BSC1 cells at an abundance similar to that of the full-length N protein. Considering this high abundance, and regardless of its origin, the 30 kDa species may have an important role in the life cycle of this virus. Since this 30 kDa species is composed of a large proportion of the PVM N protein coding sequence, it is feasible that this species may

share a domain of functional importance with the full-length N protein. If this species performs a function, it may be regulatory. Possible functional aspects of the 30 kDa N protein polypeptides are discussed in section 11.4.

CHAPTER 10

INTERACTION BETWEEN THE N AND P PROTEINS OF PVM.

10.1: THE WEST WESTERN BINDING ASSAY

The west western binding assay involves incubating *in vitro* expressed and radiolabelled proteins with proteins that are immobilised onto a membrane support. The degree to which these proteins associate can be determined by autoradiography of the washed membrane (Homann *et al.*, 1991). In negative sense, single stranded RNA viruses, the various RNA synthetic functions are carried out by the nucleocapsid complex which is formed by an association of the N, P, and L proteins. The function of the nucleocapsid complex is dependant on the presence of all three components interacting with one another and it is presumably through these interactions that the nucleocapsid complex is able to alter its function between genome replication and mRNA transcription. The mechanism responsible for this switch is still not understood, and this mainly derives from the lack of understanding of how the nucleocapsid components interact. The interactions that occur between the components of the paramyxovirus SEN have been well studied (Ryan and Portner, 1990; Ryan *et al.*, 1990; Ryan *et al.*, 1991; Homann *et al.*, 1991). However nucleocapsid component interactions have not been studied in detail for a pneumovirus of which the nucleocapsid components are clearly structurally different to those of SEN and for which the nucleocapsid interactions are also likely to be different. The west western assay was used to study the interactions that occur between the PVM N and P proteins and the results of these experiments are presented below.

10.2: THE FULL-LENGTH PVM N PROTEIN IS ABLE TO INTERACT WITH THE PVM P₃₉ PROTEIN *IN VITRO*.

By using the west western protein binding assay, ³⁵S methionine labelled rabbit reticulocyte expressed full-length PVM N protein was found to bind to a membrane immobilised PVM-infected BSC1 cell protein which had a molecular weight of 39 kDa (figure 33). Confirmation that the radioactively labelled N protein was binding to the PVM P₃₉ protein was made by using immobilised PVM P₃₉ protein expressed from *E. coli* BL21 transformed with pGEMP (figure 33). There was no binding of radiolabelled N protein to cultures of induced *E. coli* BL21 transformed with pGEM1. Radiolabelled components of the reticulocyte lysate were unable to associate with a PVM-infected cell polypeptide species.

10.3: IDENTIFICATION OF DOMAINS ON THE PVM N PROTEIN RESPONSIBLE FOR BINDING TO RENATURED P PROTEIN.

10.3.1: Introduction.

In order to identify regions of the PVM N protein that were involved in binding PVM P protein, a set of T7 transcription plasmids was constructed which were able to transcribe mRNA expressing a range of amino and carboxyl-terminally truncated and also internally deleted N-related proteins. These N-related proteins were used in the west western assay system to see if they were able to associate with immobilised P protein.

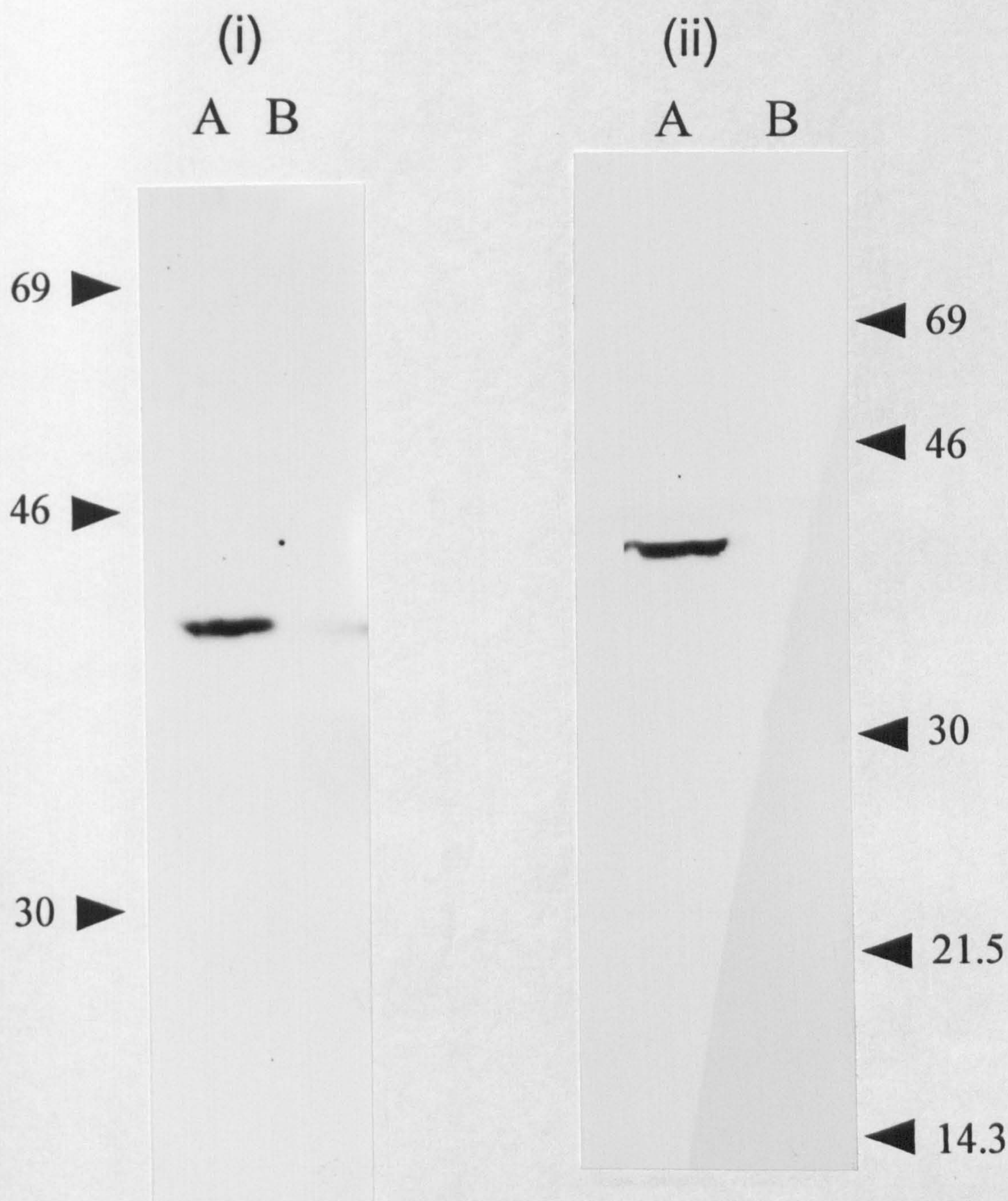


Figure 33. The full-length PVM N protein binds to full-length PVM P protein. (i). Lane 1: Radioactively labelled full-length N protein expressed from reticulocyte lysate programmed with pGEMN bound to a polypeptide with a molecular weight of 39 kDa expressed in PVM-infected cells in the west western assay. No binding occurred to a polypeptide with a similar molecular weight in mock-infected cells (lane B). (ii). Confirmation that the PVM-infected cell 39 kDa polypeptide was the PVM P protein was made by using the reticulocyte expressed labelled N protein with P protein expressed by *E. coli* BL21 transformed with pGEMP in the west western assay. Lane 1: The expression products of IPTG-induced *E. coli* BL21 transformed with pGEMP. Lane 2: Expression products of un-induced *E. coli* BL21 transformed with pGEMP. The radioactively labelled N protein binds to the bacterially expressed P protein.

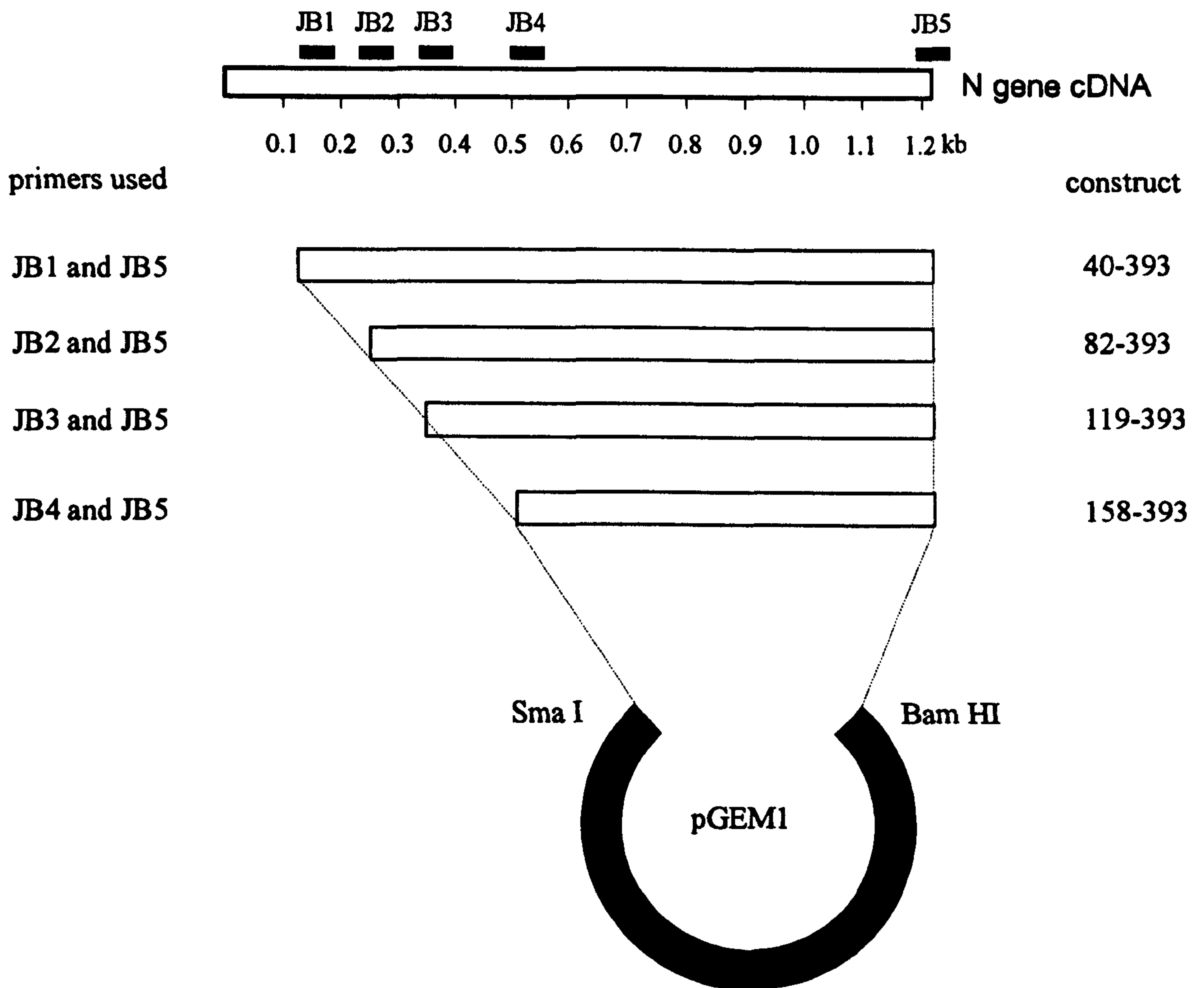


Figure 34. Construction of transcription plasmids 40-393, 82-393, 119-393 and 158-393. Regions of the PVM N gene cDNA that were lacking 5' sequences were amplified using PCR and oligonucleotide primers JB1, JB2, JB3, JB4, and JB5. The positions at which these primers anneal to the N gene cDNA is shown in the figure. The amplified fragments were blunt ended with Klenow and then digested with *Bam* HI which cuts within primer JB5. The fragments were then ligated into pGEM1 that had been digested with *Sma* I and *Bam* HI. The number designated to each construct represents the amino acids of the PVM N gene that the construct has the capacity to encode.

10.3.2: Transcription plasmids constructed using N gene cDNA.

As described in section 9.1.1, full-length and 5' terminally truncated N gene cDNA was inserted into the multiple cloning site of linearised pGEM1, using standard molecular techniques, to create the transcription plasmids pGEMN and pGEMNTR respectively. The construction of plasmids N40-393, N82-393, N119-393 and N158-393 which all feature a deleted region at the 5' end of the N gene cDNA as detailed in figure 34. As there are no appropriate restriction sites in this region of the N gene cDNA to create the appropriate deletions, the deletions were created by PCR amplification of N gene cDNA between either one of four 5' oligonucleotide primers (JB1 GGAATTCATGCAGAAAGCCCTTGCAAGG, hybridises to N gene nucleotides 149 to 169; JB2 GGAATTCATGATATTAAGAGAAGCCGGC, hybridises to N gene nucleotides 274 to 292; JB3 GGAATTCATGGATGCAGCC AATTAGCTG, hybridises to N gene nucleotides 386 to 407; JB4 GGAATTCA TGATAGTTCTCTGTATTGCAGC, hybridises to N gene nucleotides 503 to 520) which determined the precise 5' end of the construct, and a 3' oligonucleotide primer (JB5, CCCGGGATCCTAAATATCATCATCAGGAGTGTC, hybridises to N gene nucleotides 1189 to 1215) which was common to all four constructs. The location that these primers anneal to the PVM N gene cDNA is shown in figure 34. The resulting PCR generated fragments were blunt ended and cleaved at sites within the oligonucleotide primers by digestion with the restriction enzyme *Bam* HI. The cleaved fragments were isolated by agarose gel electrophoresis and purified by the fragment preparation technique after which the fragments were ligated into the multiple cloning site of pGEM1. The number designated to each of these four plasmids and all other T7 transcription plasmids constructed for use in this assay corresponds to the amino acids of the PVM N protein each cDNA clone was capable of encoding.

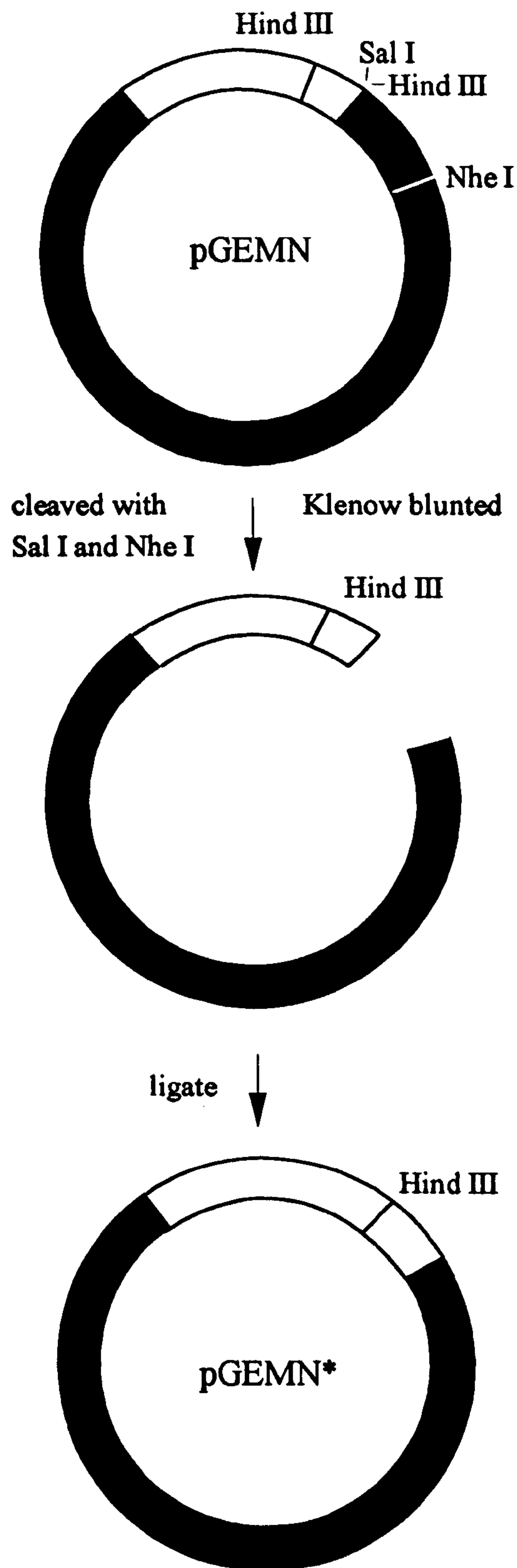


Figure 35. The removal of the *Hind* III restriction enzyme recognition site from the pGEMN multiple cloning site. pGEMN, which contains a full-length cDNA copy of the PVM N gene, was digested with restriction enzymes *Hinc* II and *Nhe* I to remove nucleotides 2623 to 24 of pGEM1. The ends of the plasmid were blunted using Klenow and ligated to form the plasmid pGEMN*. This plasmid now contained a unique *Hind* III site within the N gene cDNA segment which could be used in order to manipulate the N gene cDNA sequence.

Constructs N1-196/310-393 and N1-251/310-393 contain N gene cDNA which has had an internal segment of coding sequence removed. In order to generate Plasmids N1-196/310-393 and N1-250/310-393, the plasmid pGEMN was first modified by the removal of the *Hind* III site present within its multiple cloning region which allowed the *Hind* III site present within the N gene cDNA sequence to be utilised for the removal of an internal fragment. The construction of this modified version of pGEMN (pGEMN*) is shown in figure 35. Plasmid pGEMN was cleaved with *Hinc* II and *Nhe* I, blunt ended with Klenow and then religated. The removal of the *Hinc* II-*Nhe* I fragment was confirmed by restriction enzyme digestion. In order to derive construct N1-250/310-393, pGEMN* was then cleaved with *Nsi* I and *Hind* III which removed an internal region of the N gene cDNA sequence. The ends of the plasmid were blunted and then religated and the reading frame was maintained. Klenow is able to blunt the *Nsi* I end by digesting the protruding nucleotides and the *Hind* III end is blunted by the ability of Klenow to fill-in 5' recessed ends. Plasmid N1-196/310-393 was derived from the modified pGEMN by digestion with restriction enzymes *Kpn* I and *Hind* III. The ends of the plasmid were blunted using Klenow and the plasmid was religated. As above, Klenow is able to digest the protruding end of the *Kpn* I cleaved site and fill-in the 5' overhang on the *Hind* III cleaved site. The construction of these plasmids is shown in figure 36.

Constructs N1-249, N1-308 and N1-381 were all derived from pGEMN by digestion with restriction enzymes to remove fragments of N protein carboxyl-terminal coding region. The restriction enzymes used to remove the internal fragment in each construct are as follows. Construct N1-249 (*Nsi* I and *Bam* HI), construct N1-308 (*Hind* III), construct N1-381 (*Pvu* II and *Bam* HI). The altered plasmids were blunt ended using Klenow and then religated. Constructs N200-327 and N257-393 were derived by removal of a segment of N protein coding region from pGEMNTR, followed

by blunt-ending and re-ligation. The restriction enzymes used were *Pst* and *Bam* HI (construct N200-327), and *Eco* RI and *Nsi* I (construct N257-393).

A transcription plasmid capable of directing the expression of the HRSV N protein was also constructed by inserting a full-length cDNA copy of the HRSV N gene (a gift from Dr. C. Caravokyri, University of Warwick, UK.) into pGEM1.

The N gene cDNA fragments present in all the above constructs are schematically represented in figure 37.

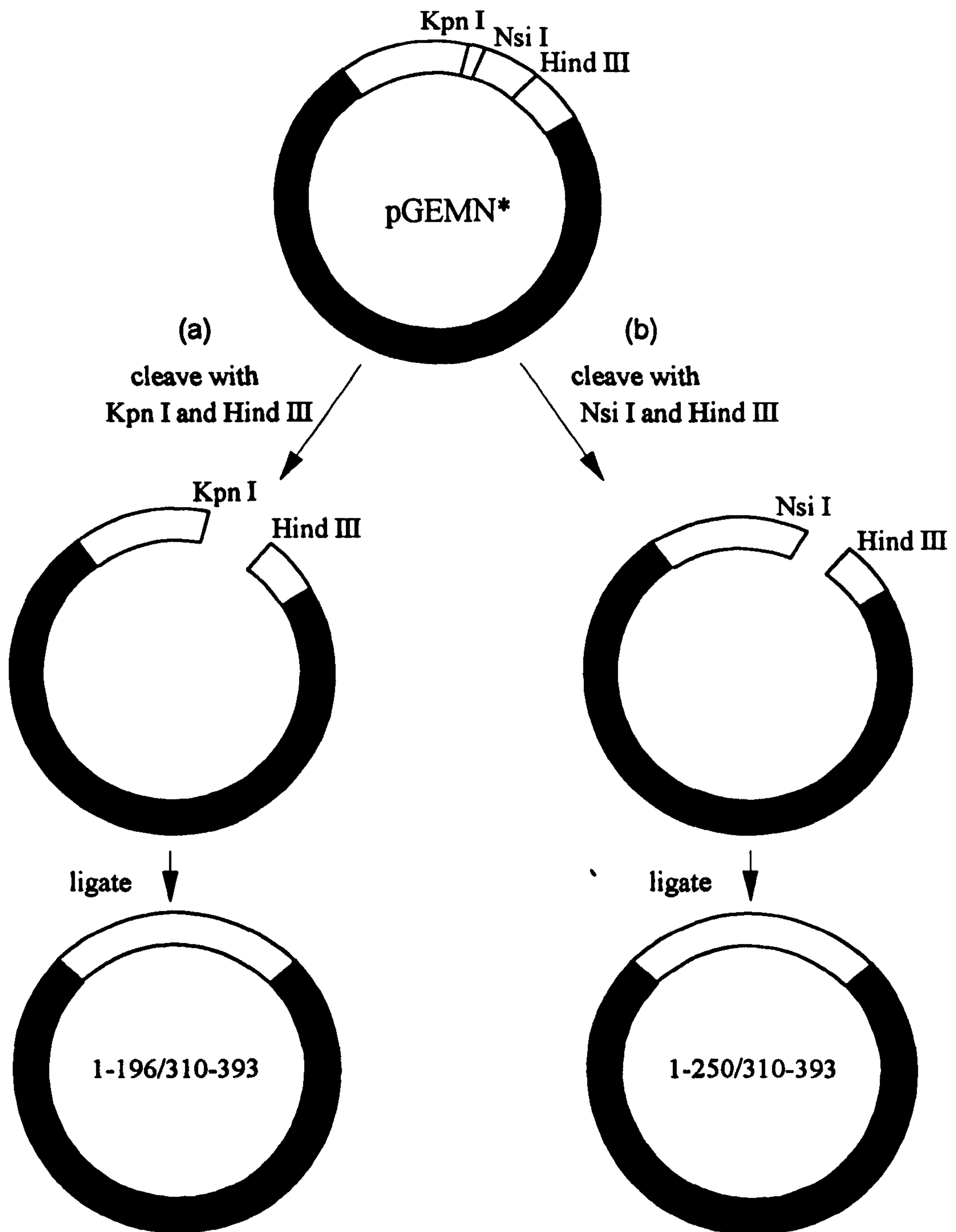


Figure 36. Construction of transcription plasmids 1-196/310-393 and 1-250/310-393 which contain N gene cDNA fragments which have internally deleted regions. (a). pGEMN* (construction details figure 36) was cleaved with restriction enzymes *Kpn* I and *Hind* III which removed nucleotides 621-954 of the N gene cDNA. The ends of the plasmid were blunted using Klenow and then religated to create the plasmid 1-196/310-393. The reading frame was maintained. **(b).** pGEMN* was cleaved with restriction enzymes *Nsi* I and *Hind* III which removed nucleotides 777-954 of the N gene cDNA. The ends were blunted using Klenow and then religated to form plasmid 1-250/310-393. The reading frame was maintained.

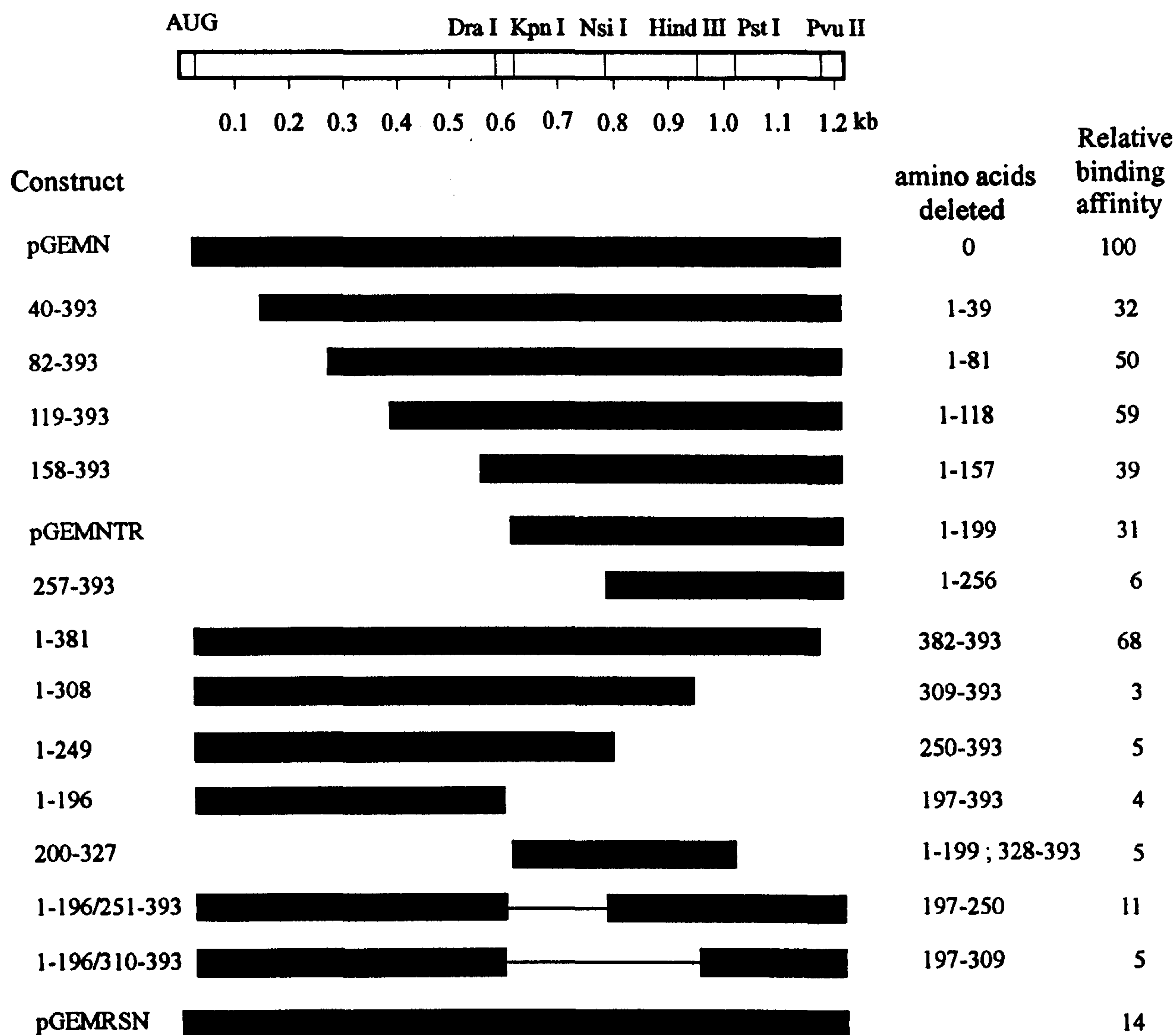


Figure 37. Diagrammatical representation of the N gene cDNA fragments contained within the T7 transcription plasmids used to express the truncated and internally deleted N protein species for use in the west western binding assay. The PVM N gene is represented diagrammatically at the top of the figure, and the positions of the restriction enzyme recognition sites that were used for the construction of these transcription plasmids are indicated. The number of the construct that contained each individual N gene cDNA fragment is given at the 5' end of each fragment. The designated number details the amino acids of the N protein that each cDNA fragment encodes. The amino acids of the PVM N protein that are deleted in the subsequently expressed N protein species is shown on the 3' end of each fragment. The ability of the N-related proteins to bind to the PVM P protein in the west-western assay is also shown, and is expressed as a percentage of the ability of full-length N protein to bind P protein.

10.3.3: *In vitro* expression of N protein related polypeptides.

From this set of T7 transcription plasmids, mRNAs were transcribed which were able to direct the synthesis of a series of internally deleted and amino- and carboxyl-terminally truncated radioactively labelled forms of the N protein in a rabbit reticulocyte lysate system. SDS-PAGE analysis of the altered radiolabelled N protein species revealed that all the species migrated to give an observed molecular weight which was consistent with their predicted molecular weights (figures 38 and 39). In addition to the presence of the intended altered N-related proteins, minor species of a lower molecular weight were commonly seen as expression products of the various mRNA species. These are expressed as a result of initiation events on internal AUG initiation codons located downstream of the first and therefore most frequently used initiation codon present on each of the mRNA species. The initiation codon located at nucleotide positions 629 to 631 on the N gene is particularly favourable as a site of internal initiation due to its favourable context (section 9.3.2), and initiation seems to have occurred from this site during the expression of many of the constructs. It is likely that the internally initiated N-related proteins will probably not be as effective at binding P protein than the intended full-length product of each altered mRNA because, by the nature of their expression, they will represent only a portion of the intended full-length species. The inclusion into the binding assay of a quantity of polypeptides that are less able to bind immobilised P protein than the intended species may lower the apparent P protein binding ability of this species. This is because each set of translation products was applied to the membrane in a quantity that was determined by the level of incorporated radioactivity and so if an amount of poorly binding species were included in this quantity then there would be a lower proportion of the intended species present. However since in the case of each altered N-related protein

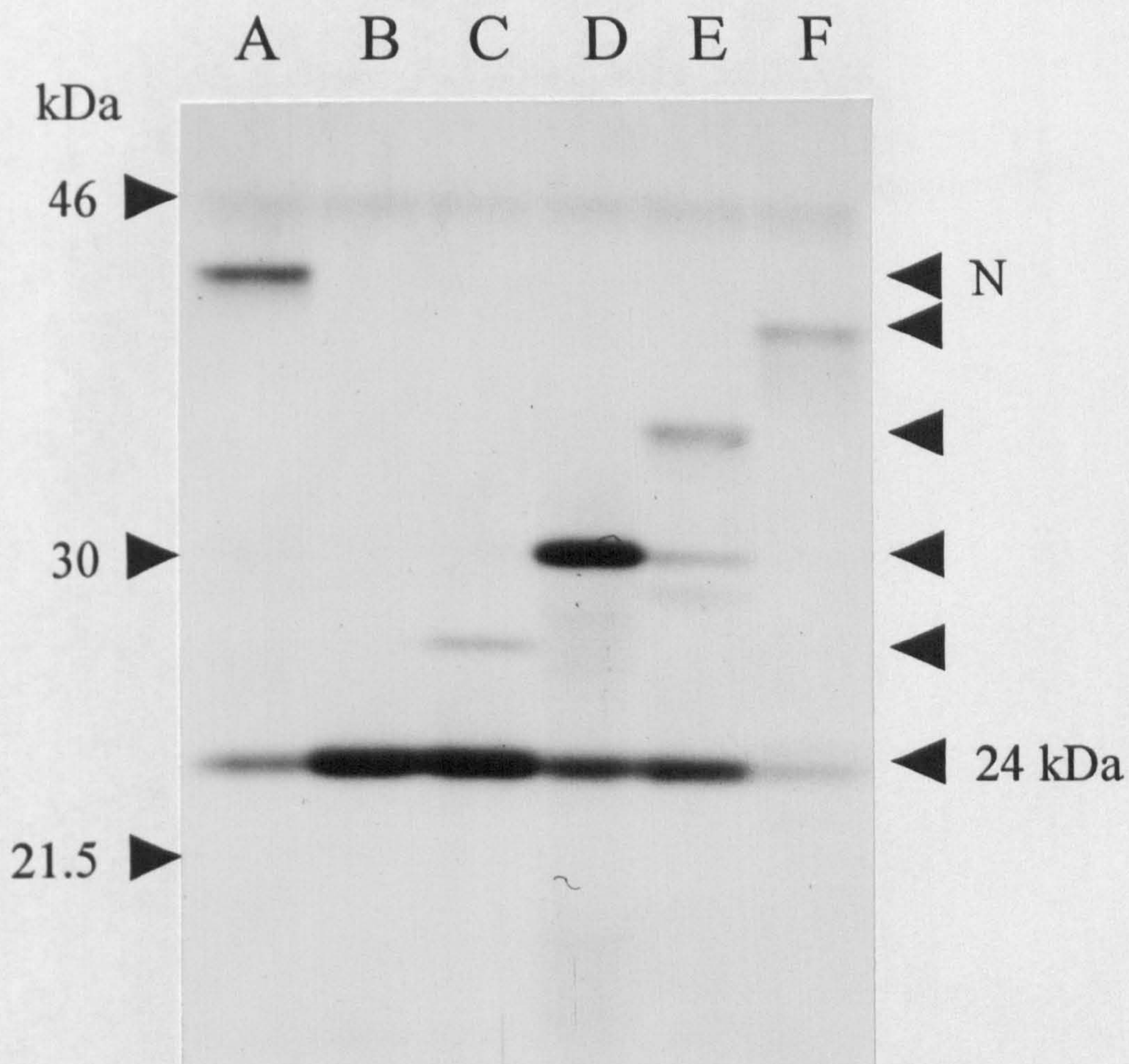


Figure 38. Expression of truncated PVM N proteins for use in the west western binding assay. mRNAs were transcribed from a set of T7 transcription plasmids which contained inserts capable of directing the synthesis of amino-terminally truncated versions of the PVM N protein. The mRNAs were used to programme a reticulocyte lysate system and the radioactively labelled expression products were analysed by SDS-PAGE and the resulting gel was autoradiographed in order to visualise the polypeptides. The plasmids used to transcribe the mRNAs were as follows. pGEMN (lane A), pGEMNTR (lane B), N40-393 (lane F), N82-393 (lane E), N119-393 (lane D), N158-393 (lane C). Construction of these plasmids is detailed in section 10.3.2. The position of molecular weight markers are marked on the left.

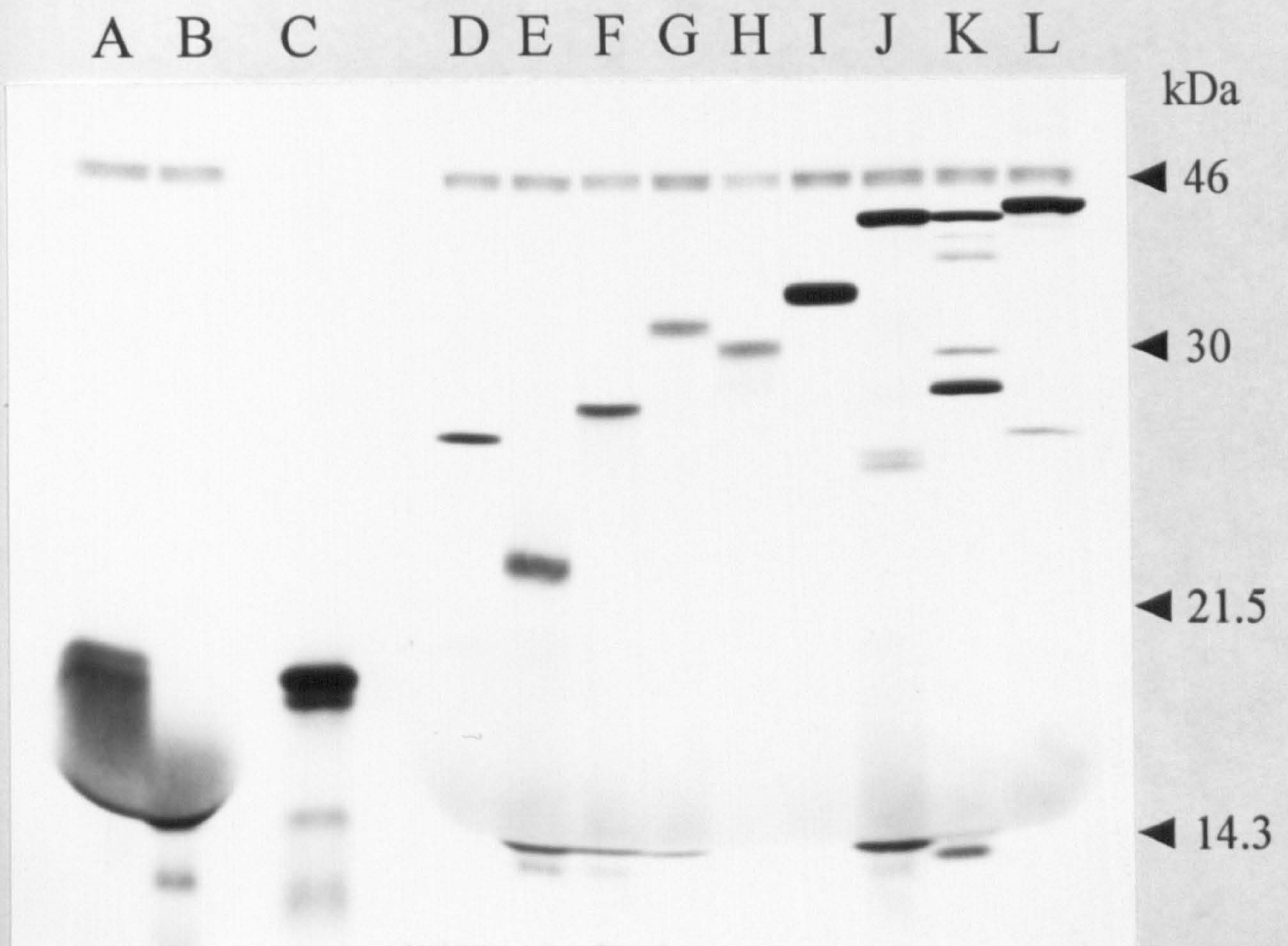


Figure 39. Expression of modified PVM N proteins for use in the west western binding assay. mRNAs were transcribed from a set of T7 transcription plasmids which contained inserts capable of directing the synthesis of amino-terminally, carboxyl-terminally and internally deleted versions of the PVM N protein. The mRNAs were used to programme an *in vitro* translation system and the radioactively labelled expression products were analysed by SDS-PAGE. The resulting gel was autoradiographed in order to visualise the polypeptides. The plasmids used to transcribe the mRNAs were as follows. N257-393 (lanes A and C), N200-327 (lane B), pGEMNTR (lane D), N1-196 (lane E), N1-249 (lane F), N1-308 (lane G), N1-196/310-393 (lane H), N1-196/251-393 (lane I), N1-381 (lane J), pGEMRSN (contains full-length HRSV N gene cDNA, lane K), pGEMN (lane L). With the exception of lane C, for which a wheat germ lysate was used, all translations were performed in reticulocyte lysate. The construction of the above plasmids is detailed in section 10.3.2. The position of molecular weight markers is indicated on the right.

construct, the intended polypeptide is present as the overwhelmingly major species, the minor internally initiated species probably do not make a significant difference to the calculated P protein binding affinities.

10.3.4: The ability of the altered N protein species to bind to PVM P protein.

By using the binding assay, it was possible to test each of the ^{35}S methionine labelled N protein species for their ability to bind to PVM P protein present in the PVM-infected BSC1 cell material immobilised onto the PVDF membrane. Autoradiographs of the filters on which the binding assays were performed are shown in figures 40 and 41. The autoradiographs were subjected to densitometric analysis to quantify the extent of binding that had occurred. The optical density of the autoradiograph that corresponded to an area of the filter immediately adjacent to a region where binding had occurred was taken as background and was subtracted from the optical density reading of the autoradiograph region which corresponded to the bound protein. In order to compare the binding abilities of each of the N-related proteins, the optical density of the autoradiograph band representing bound N-related proteins had to be related to the number of molecules of the radioactively labelled species bound to the immobilised P protein at that position. The N-related proteins that were applied to the membrane contained different numbers of radiolabelled methionine residues and consequently to relate the bound N-related protein species signal intensity to an actual number of molecules of that species, an adjustment factor was introduced which, for example, would account for the increased signal that a full-length N protein molecule, which contained nine radioactively labelled methionines, would produce compared with a truncated N-related proteins containing only three labelled methionines. The number of molecules of each N protein species applied to the membranes were not accurately standardised although the 30,000 cpm of incorporated label initially incubated with the membrane would contain

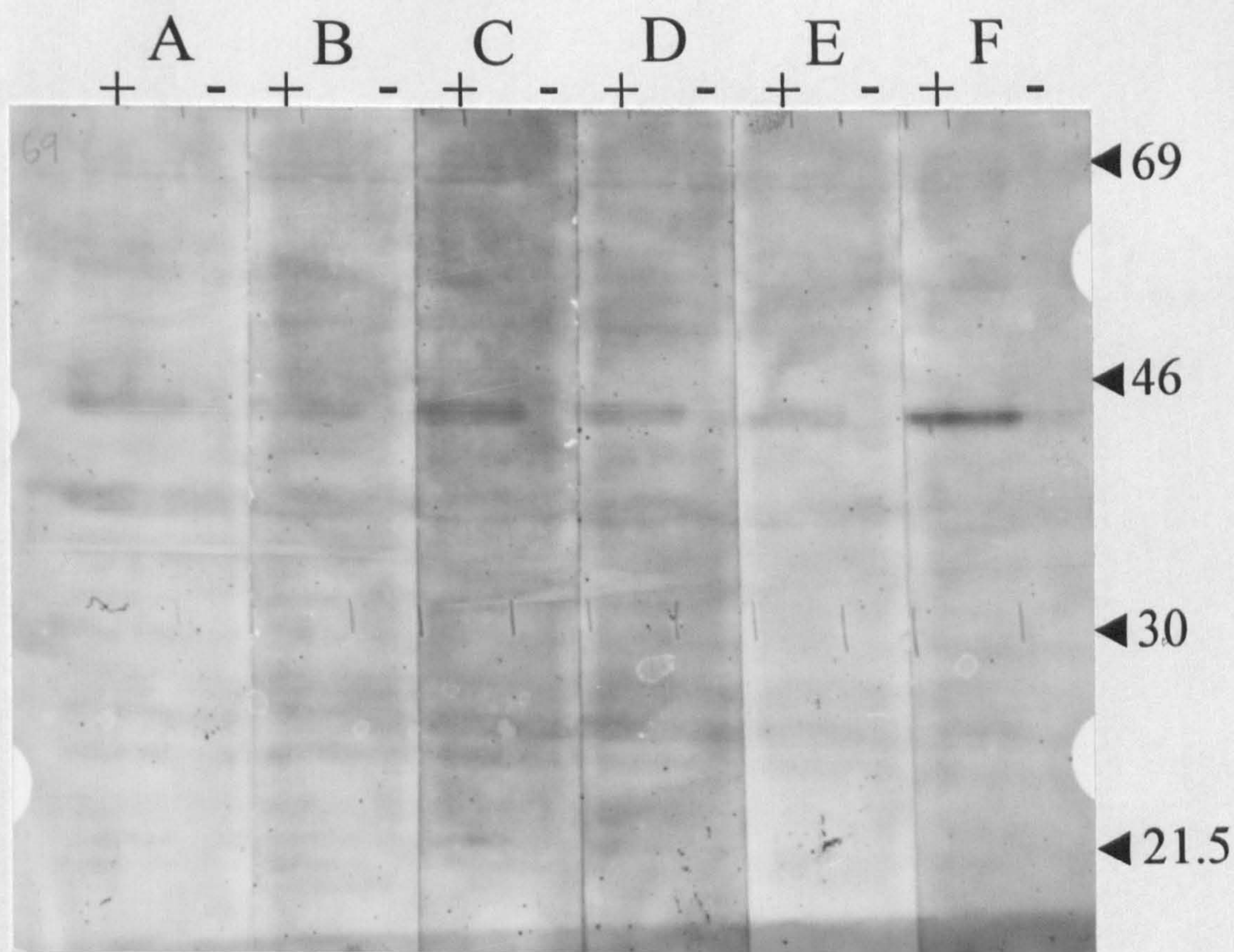


Figure 40. Determination of the regions of the PVM N protein that are required in order to bind PVM P protein. A series of modified N proteins were expressed using reticulocyte lysate programmed with mRNAs transcribed from T7 transcription vectors containing PCR amplified N gene cDNA fragments. The construction of these plasmids is detailed in section 10.3.2, and the polypeptides expressed by these plasmids are shown on figure 38. The radioactively labelled altered N proteins were then used in the west western assay with PVM-infected and mock infected cell material that had been subjected to SDS-PAGE and immobilised onto a PVDF membrane. The membrane was cut into strips of which the left hand lane (+) contained PVM-infected cell material and the right hand lane (-) contained mock-infected cell material. The plasmids used to express the various N proteins are as follows. N40-393 (strip A), N82-393 (strip B), N119-393 (strip C), N158-393 (strip D), pGEMNTR (strip E), pGEMN (strip F). The position of molecular weight markers is shown on the right.

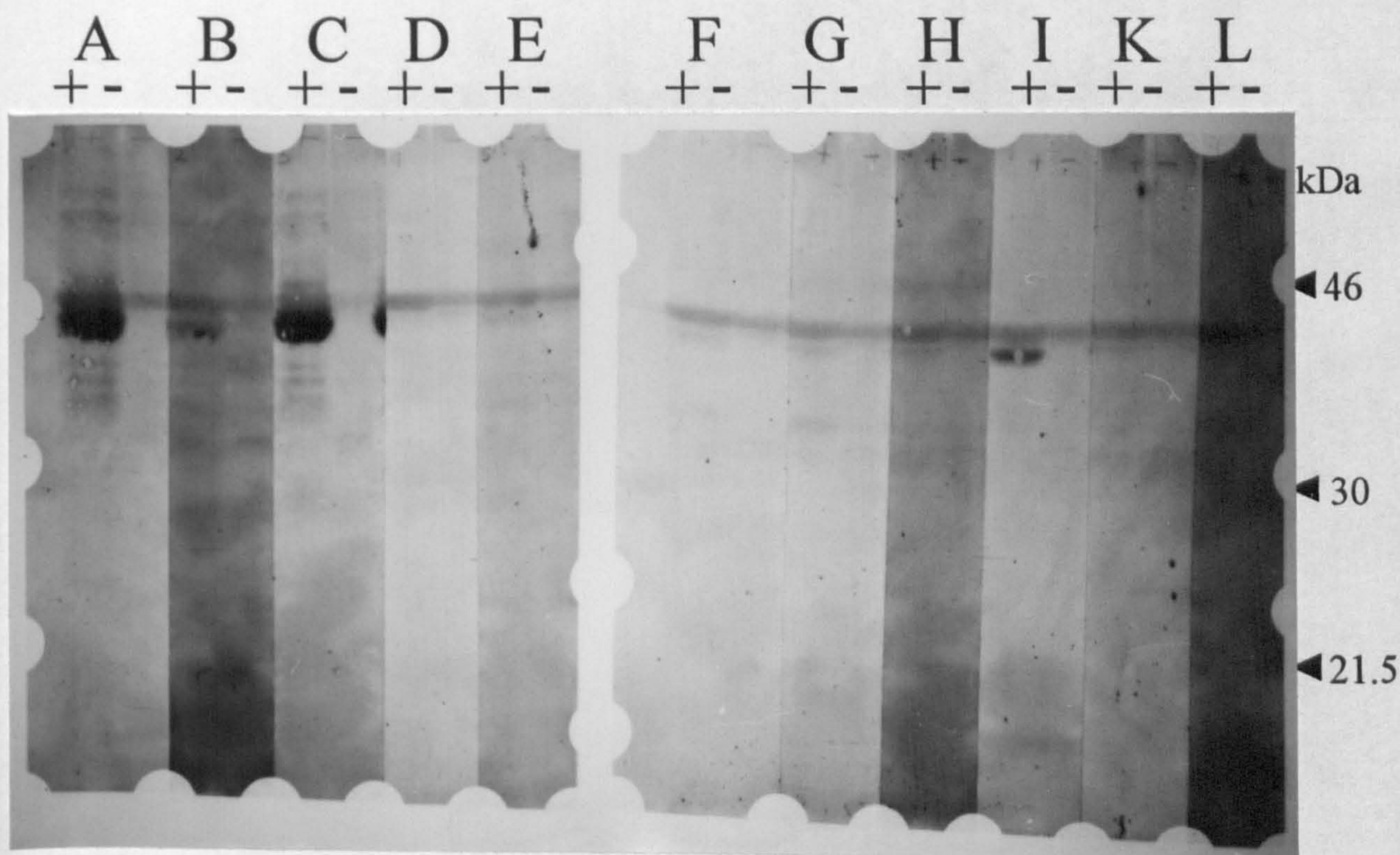


Figure 41. Determination of the regions of the PVM N protein that are required in order to bind PVM P protein. A series of modified N proteins were expressed using reticulocyte lysate programmed with mRNA transcribed from T7 transcription vectors containing altered N gene cDNA fragments. The construction of these plasmids is detailed in section 10.3.2, and the polypeptides expressed by these plasmids are shown on figures 38 and 39. The radiolabelled altered N proteins were then used in the west western assay with PVM infected and mock-infected cell material that had been subjected to SDS-PAGE and then immobilised onto a PVDF membrane. The membrane was cut into strips of which the left hand lane (+) contained PVM infected cell material and the right hand lane (-) contained mock infected cell material. The plasmids used to express the various N proteins are as follows pGEMN (strip A), pGEMRSN (contains full-length HRSV N gene cDNA, strip B), N1-381 (strip C), N1-196/251-393 (strip D), N1-196/310-393 (strip E), N1-308 (strip F), N1-249 (strip G), N1-196 (strip H), pGEMNTR (strip I), N200-327 (strip J), N257-393 (strip K). The position of molecular weight markers is indicated on the right.

approximately equal numbers of radioactively labelled N protein species. However as the P protein would be present in vast excess, sub-saturation kinetics would apply to the association reaction and consequently, the small variation in the initial quantity of applied N protein species would not affect the degree of interaction between the N and P proteins. The extent to which radiolabelled full-length N protein bound to membrane immobilised P protein was considered to be 100%, and the binding of all other radiolabelled N-related proteins was expressed relative to this. These binding affinity values are listed next to the corresponding construct in figure 37. The values represent the means of binding affinities calculated from two identical binding experiments. The autoradiograph obtained from only one of these experiments is shown in figures 40 and 41.

All the amino acid deletions introduced into the various expressed truncated N proteins reduced their relative P protein binding affinities. These deletions spanned most of the coding region of the N protein and thus it seems that large parts or several domains of the N protein are involved in the binding process. The expression product of construct pGEMNTR which represents the carboxyl-terminal half of the full-length N protein was able to bind the P₃₉ protein with a relative binding affinity of 18% whereas the expression product of construct N1-196 which comprises the amino-terminal half of the full-length N protein bound P protein with an affinity of only 4%. Although this figure of 4% is small, specific binding was still clearly taking place and may reflect an inherent ability of certain portions of the N protein to contribute to the overall binding of the full-length protein. Polypeptides encoded by constructs N1-249, N1-308 and N1-381 were all carboxyl-terminally extended derivatives over the product of construct N1-196. Polypeptides encoded for by constructs N1-249 and N1-308 display approximately equal but extremely low P protein binding affinities, whereas the polypeptide generated from construct N1-381, which lacks only 12 carboxyl-terminal amino acids of the N protein, has a relative binding affinity of 68% over full-length N

protein. This observation suggests that the immediate carboxyl-terminal region of the N protein is involved in the P protein binding process, but that it is not essential for it. Constructs N257-393 and N200-327 are derived from construct pGEMNTR by removal of either the amino-terminal or carboxyl-terminal N protein coding region respectively. Polypeptide products expressed from both constructs displayed considerably reduced binding affinities (6% and 5% respectively), when compared to the product of the original construct pGEMNTR. Expression products of constructs N1-196/251-393 and N1-196/310-393 have lost internal regions of N protein coding sequence, and both show a marked decrease in P protein binding affinity relative to full-length N protein (11% and 5% respectively). The N-related proteins expressed from the PCR generated constructs N40-393, N82-393, N119-393 and N158-393 are all related to one another by progressive truncations at their amino-terminal ends. The largest of these truncated N proteins, expressed from construct N40-393, has lost only 39 residues from the amino-terminal end of the full-length N protein and has a relative binding affinity of 32%. It thus seems that the immediate amino-terminal region of the N protein is essential for efficient binding of P protein. The expression products of constructs N82-393, N119-393, N158-393, pGEMNTR and N257-393 display respective binding affinities of 50%, 59%, 39%, 31% and 6%. This series of N protein species, having progressively truncated carboxyl-terminal regions, displays approximately progressive reduction in P protein binding affinity.

The relative binding affinity of the HRSV N protein for the PVM P protein is a reflection on the degree of similarity that exists between the N proteins of PVM and HRSV.

10.4: IDENTIFICATION OF DOMAINS ON THE PVM P PROTEIN THAT ARE INVOLVED IN BINDING TO PVM N PROTEIN.

10.4.1: Introduction.

Binding of ^{35}S methionine labelled P protein species to membrane immobilised N protein was not detected using the west western assay described above. In order to identify domains on the PVM P protein involved in binding N protein an alternative approach was adopted which involved immobilising bacterially expressed amino-terminally and carboxyl-terminally truncated versions of the PVM P protein to a membrane support and determining which of these altered species were able to retain radioactively labelled full-length N protein. The carboxyl-terminally truncated P proteins were expressed using the transcription plasmids described in section 10.4.2 and the amino-terminally truncated P proteins were expressed using transcription plasmids pGEMP and pGEMP37 which are described in sections 4.1 and 4.2.2 respectively.

10.4.2: Construction of plasmids to express carboxyl-terminated P protein species.

A range of plasmids were constructed which were capable of directing the expression of various carboxyl-terminally truncated versions of the PVM P protein in *E. coli* BL21. Construction details of these plasmids is shown in figure 42. The full-length P gene cDNA insert within pGEM1 was cleaved with a variety of restriction enzymes to remove regions of the P protein coding sequence. The restriction enzymes used in the constructions are shown in figure 42. The ends of the plasmid were blunt ended using Klenow and then re-ligated. In this way the transcription vectors pGEMP S_{ty} , pGEMP H_{inc} and pGEMP S_{m} were constructed. Figure 42. Construction of pGEMP S_{ty} , pGEMP S_{tu} and pGEMP H_{inc} .

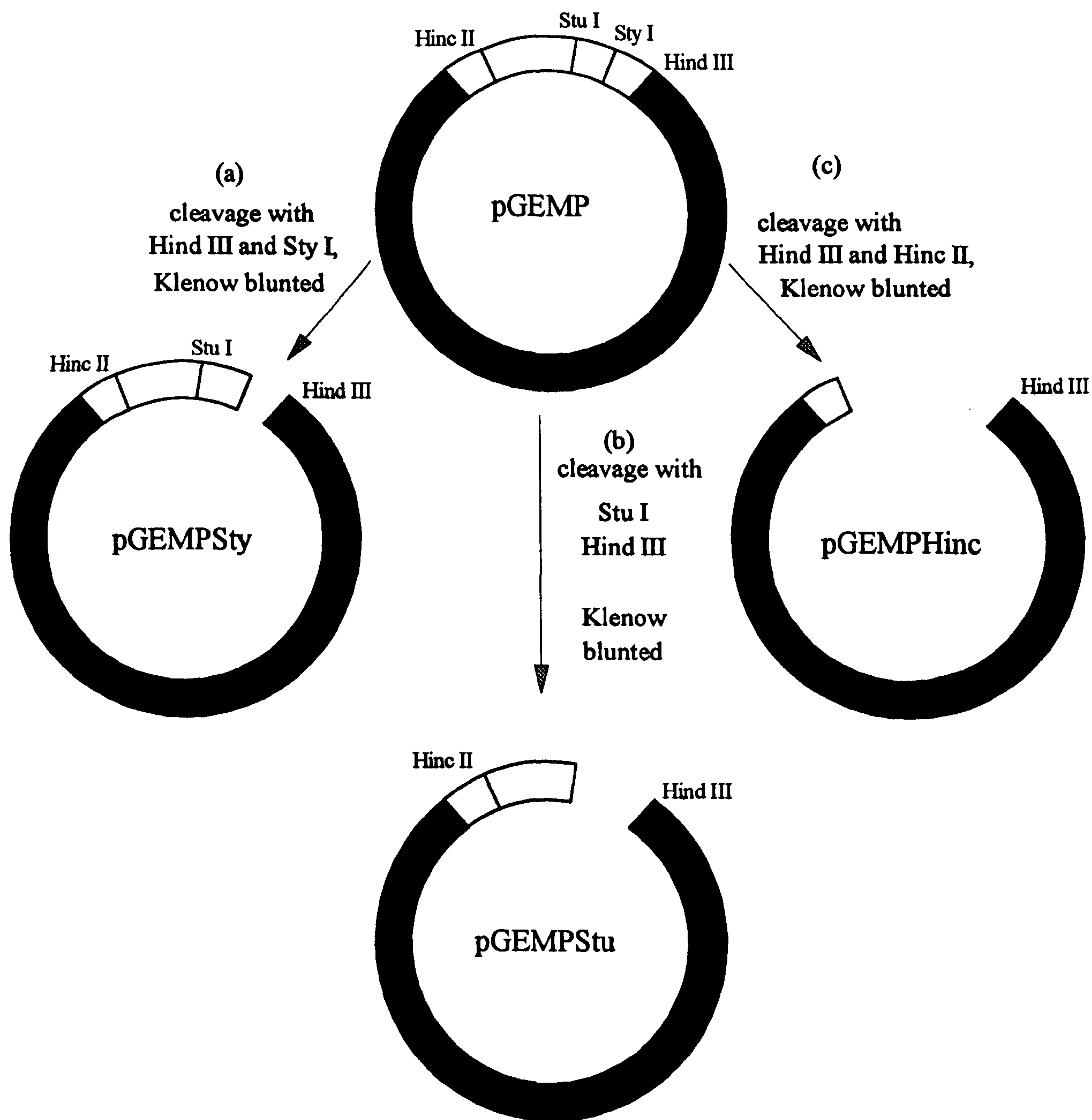


Figure 42. Construction of transcription plasmids pGEMPSty, pGEMPStu and pGEMPHinc. (a) The full-length P gene cDNA fragment within pGEMP was cleaved by restriction enzymes *Sty* I and *Hind* III which removed a segment containing P gene nucleotides 731-903. The resulting ends of the cleaved pGEMP were blunted using Klenow and religated to form the plasmid pGEMPSty. **(b).** pGEMP was cleaved with *Stu* I and *Hind* III to remove P gene nucleotides 591-903. The ends were blunted and then religated to form the plasmid pGEMPStu. **(c).** pGEMP was digested using *Hinc* II and *Hind* III to remove P gene nucleotides 289-903. The ends were blunted and religated to create plasmid pGEMPHinc.

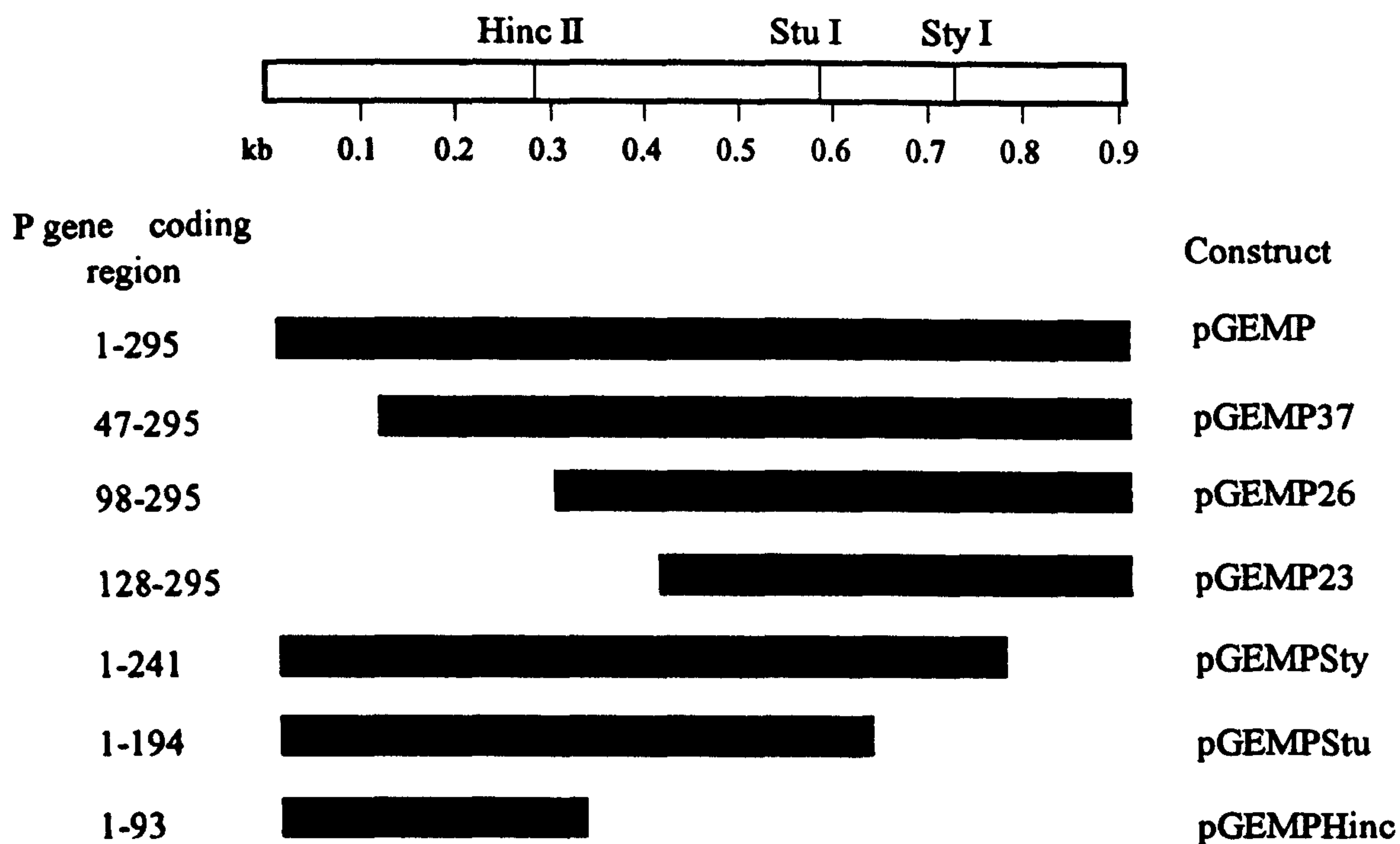


Figure 43. Diagrammatical representation of the PVM P gene cDNA fragments contained within the T7 transcription plasmids used to express the truncated and internally deleted P protein species for use in the west western binding assay. The PVM P gene is represented diagrammatically at the top of the figure, and the positions of the restriction enzyme recognition sites that were used for the construction of these transcription plasmids are indicated. The number designated to each constructs is given on the right of each fragment. The amino acids that are present in the P protein related polypeptides encoded for by the P gene cDNA fragments is shown at the left of each fragment. Only the full-length P protein (amino acids 1-295) was able to bind to N protein using the west western assay

The regions of the P gene that were used to express altered P protein species are diagrammatically represented in figure 43. These transcription plasmids were transformed into *E. coli* BL21 in a method identical to that used for the transformation of *E. coli* TG2 (section 2.5.2).

10.4.3: The abilities of the altered P protein species to bind to PVM N protein.

By expressing the amino-terminally and carboxyl-terminally truncated P protein species in *E. coli* BL21, quantification and accurate standardisation of the amount of the modified P proteins initially immobilised to the membrane was not possible. However if the immobilised P protein species were present in excess over the radiolabelled N protein, which was probably the case, the kinetics of the reaction would suggest that standardisation of the quantity of P protein species immobilised onto the membrane would have little effect on the outcome of the binding experiments. Despite this, an attempt was made to immobilise each of the P protein species to the filter in approximately equimolar amounts. This was achieved by estimating the abundance of each of the P protein species in the *E. coli* BL21 expression products from a Western blot (figure 44) and then proportionately scaling up the quantities of bacterial expression product that were applied to the west western filter. The filter was then incubated with ³⁵S methionine labelled N protein synthesised *in vitro* using pGEMN. This construct also directed the synthesis of the 24 kDa N protein related polypeptide discussed in section 9.1.2. An autoradiograph of the filter on which the binding assay was performed is shown in figure 45.

None of the amino-terminally deleted P protein species expressed from pGEMP37 were able to bind to radiolabelled full-length N protein. Similarly, none of the polypeptides expressed from the plasmids pGEMP*Sty*, pGEMP*Stu* or pGEMP*Hinc* were able to bind the *in vitro* expressed full-length N protein. It thus

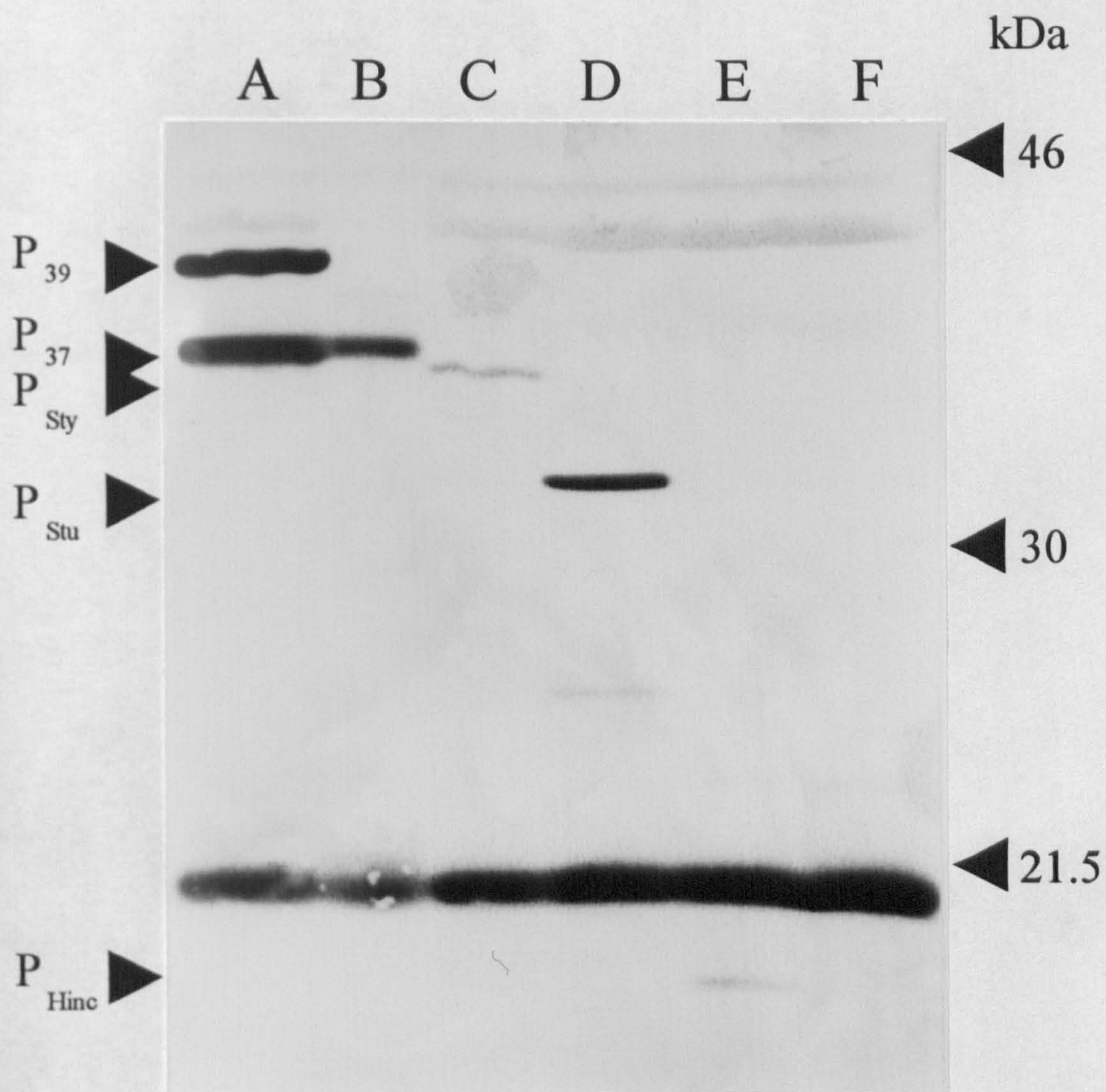


Figure 44. Western blot of truncated P proteins used in the west western blot assay. The expression products of IPTG-induced *E. coli* BL21 transformed with transcription plasmids pGEMP (lane A), pGEMP37 (lane B), pGEMPSty (lane C), pGEMPStu (lane D), pGEMPHinc (lane E) and pGEM1 (lane F) were subjected SDS-PAGE analysis, and then Western blotted using monoclonal antibodies 26/3/B5 and 26/1/A2. The positions of the polypeptides encoded for by these transcription plasmids are marked on the left. The position of molecular weight markers are indicated on the right.

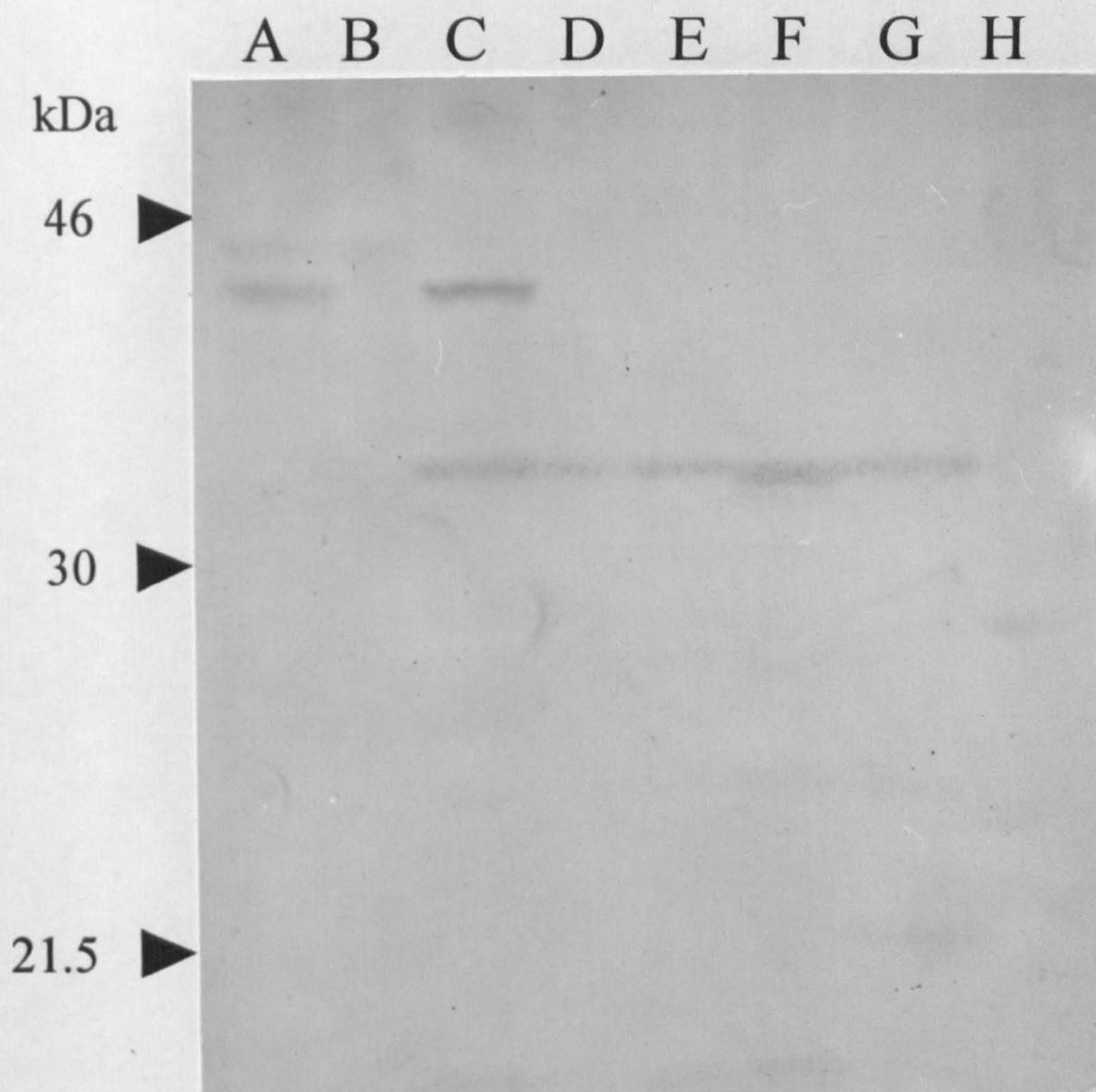


Figure 45. Determination of the regions of the PVM P protein that are required in order to bind PVM N protein. *E. coli* BL21 was transformed with a set of T7 transcription plasmids containing various P gene cDNA inserts that were capable of directing the expression of amino and carboxyl terminally truncated P proteins. The IPTG-induced bacterial expression products were immobilised onto a PVDF membrane and used with radioactively labelled full-length N protein in the west western assay. For comparison, PVM-infected (lane A) and mock-infected (lane B) cells were also immobilised onto the membrane. The transcription plasmids used to transform the *E. coli* BL21 are as follows. pGEMP (lane C), pGEMP37 (lane D), pGEMPSty (lane E), pGEMPStu (lane F), pGEMPHinc (lane G) and pGEM1 (lane H). The position of molecular weight markers are shown on the left.

seems that regions at both the amino-terminus and the carboxyl-terminus of the P protein are required to be present for the PVM P protein to bind to N protein.

10.5: DISCUSSION.

10.5.1: PVM N protein binds to immobilised PVM P protein.

Under the conditions of this assay, the PVM N protein is able to specifically interact with the PVM P protein. In order to identify regions of the N protein important in binding to P protein, a series of truncated and internally deleted forms of the N protein were expressed and used in the assay to see if they were capable of binding to the P protein present in immobilised PVM-infected BSC1 cell material. To find regions of P protein involved in binding to N protein, several bacterially expressed and modified forms of the P protein were immobilised onto a membrane and were assayed for their ability to bind radiolabelled full-length N protein.

10.5.2: The regions of the N protein that are involved in binding P protein.

Analysis of the interaction that was observed between P protein and the deleted N proteins revealed that the regions involved in binding to P protein are spread throughout the length of the N protein. All of the deletions introduced into the N protein coding region caused some degree of reduction in the ability to bind P protein. This observation does not suggest that all the residues in the N protein coding region are directly involved in forming an association with the P protein. It is possible that the ability of N protein to bind P protein is dependent on maintenance of a three-dimensional structure which holds certain domains of the N protein in the correct conformation for binding to occur. Removal of amino acids from any part of the N protein probably has the effect of altering this three-dimensional structure and

consequently the binding domains will no longer be arranged in the correct conformation for optimal binding to occur. If the amino acids that are removed are from the binding domains themselves then binding activity will be significantly reduced. Another possibility is that the N protein contains several P protein binding domains, some of which are able to bind P protein independently, and these domains all contribute to the binding of P protein in a cooperative manner. Although deletions did cause a reduction in binding affinity, in no construct was binding totally abolished. This suggests that all residues that were deleted in the constructs do contribute to the overall P protein binding capability of the N protein. This observation is similar to the findings of Homann *et al.*, (1991) which suggested that large regions of the SEN N protein were involved in binding to SEN P protein. The polypeptide products expressed from construct N1-196 and construct pGEMNTR, which represent different halves of the N protein, are able to bind P protein with affinities of 4% and 31% respectively even though they do not possess any common sequence. This suggests that there are independent domains in both the amino and carboxyl-terminal halves of the N protein which are on their own capable of binding P protein, albeit with low affinity. The expression product of construct pGEMNTR was able to bind to P protein with an affinity over 7 times that of the expression product of construct N1-196 which suggests that the carboxyl-terminal binding domain may contribute more to the binding of P protein than the amino-terminal domain. The N protein species encoded for by construct N1-382 has lost only the carboxyl-terminal 11 residues of the full-length N protein. This deletion has reduced the P protein binding affinity of this species to 68% of the full-length N protein which indicates that the extreme carboxyl-terminus of the N protein is involved to some extent but is not a major determinant in the process of binding P protein.

The region of the PVM N protein spanning residues 220 to 315 displays extremely high homology with the corresponding region of the HRSV N protein (Barr *et al.*, 1991). The conservation of amino acid sequence that is seen here suggests that

the homologous region may contain a functionally important site. The polypeptides expressed from construct N1-196/251-393 has lost amino acids 197 to 250 which are partly from this region and this deletion resulted in a P protein binding affinity of only 11%. Perhaps the homologous region represents an area of the PVM N protein where conformation is particularly sensitive to structural changes such that removal of any part of this region severely reduces P protein binding ability. The 54 amino acids deleted from the protein expressed from construct N1-196/251-393 are unlikely to constitute an independent P protein binding domain since the expression product of construct N200-327 contains this region plus an additional 83 carboxyl-terminal residues, but displays a P protein binding affinity of only 5%. This further reinforces the suggestion that the conformation of the N protein is important in determining binding capacity. It is not known what structural feature(s) of the PVM N protein makes it able to bind P protein. The P protein is highly negatively charged, particularly at the carboxyl-terminus, and so maybe the interaction that occurs is mediated by a binding domain on the PVM N protein that is positively charged.

10.5.3: Regions of the P protein that are involved in binding N protein.

The regions of the P protein involved in binding to N protein were investigated using a limited range of deleted P protein constructs. None of the amino-terminally truncated P protein species immobilised on the membrane were capable of binding to the radiolabelled, full-length N protein. The observation that the *E. coli* expressed P₂₆ or P₂₃ polypeptides were not able to bind ³⁵S methionine labelled N protein was not surprising as labelled N protein did not bind to the 26 kDa and 23 kDa P-related polypeptides expressed from PVM-infected BSC1 cells (section 10.2). The 37 kDa P protein species expressed from pGEMP37 has lost only 40 amino-terminal amino acids of the full-length P protein and so the inability of this species to bind N protein indicates that the extreme amino-terminus of the P protein is essential for maintenance

of N protein binding ability. Additionally, none of the carboxyl-terminally truncated P protein species were capable of binding to radiolabelled N protein. The P protein species expressed from pGEMPS_{ty} has lost only the carboxyl-terminal 57 amino acids and since the loss of these residues coincided with the loss of N protein binding ability, the extreme carboxyl-terminal region of P protein also appears to be essential for preservation of N protein binding activity. Neither region however comprises a complete N protein binding site since the amino-terminal 40 amino acids is contained in the expression product of pGEMPS_{ty} and conversely the carboxyl-terminal 57 amino acids are contained within the expression product of pGEMP37, and neither species can bind N protein. These results suggest that regions at both the amino-terminus and also the carboxyl-terminus are needed for the formation of a functional N protein binding site. The finding that both of the extreme termini of the PVM P protein are indispensable for N protein binding to occur is in contrast to the results obtained for the SEN P protein in which removal of the amino-terminal 344 amino acids had no affect on its ability to bind N protein (Ryan *et al.*, 1991). This observation indicates that the pneumovirus P protein may have a different arrangement of functional domains to the P proteins of other members of the family *Paramyxoviridae*.

The ability of bacterially expressed P protein to associate with PVM N protein in this assay suggests that phosphorylation may not be important in the binding process, since bacterial systems are not considered able to be capable of the appropriate phosphorylation that is a feature of mammalian tissue culture systems (Barik and Banerjee, 1991).

By analogy with other closely related viruses, including HRSV, the PVM N protein would be expected to be able to bind to itself (Samal *et al.*, 1993). However, using this assay, binding of radiolabelled N protein to immobilised N protein was not observed. Radiolabelled N protein did not bind to any PVM-infected BSC1 cell species other than PVM P protein.

Since radiolabelled N protein was able to bind immobilised P protein, the ^{35}S methionine labelled P protein might have been expected to bind to N protein present in the immobilised PVM-infected BSC1 cell material. However, at no time in this study was immobilised N protein observed to bind either reticulocyte expressed N protein or reticulocyte expressed P protein. It is possible that the renaturation conditions used for the immobilised PVM-infected BSC1 cell material were sufficient to restore the binding capabilities of immobilised P protein such that it could bind to radiolabelled N protein, but conditions were not suitable to restore the binding ability of the immobilised N protein for both radiolabelled N and P protein species. It thus seems that the immobilised N and P protein species require different conditions to allow restoration of their respective binding capabilities. These observations also suggest that the insufficient renaturation that the immobilised N protein has experienced has affected not only the domain that binds the P protein but also the domain responsible for interaction with other N protein molecules.

By the use of the PCR, more highly resolved N and P protein expressing constructs may be generated which, in conjunction with the binding assay described here, would provide further information as to which residues are required for interaction with nucleocapsid components. The assay system used in this study was able to investigate the way in which two virally encoded proteins interacted with each other in isolation from other viral proteins. However, the pneumovirus nucleocapsid complex is, by analogy with the paramyxovirus complex, thought to be composed of 4 components, the N protein, the P protein, the L protein and the RNA template. It is possible that the interaction that occurs between just 2 components of the nucleocapsid complex, as has been investigated in this study, may differ if other nucleocapsid complex components were present. Another important point to bear in mind in the interpretation of these binding assay results is that *in vivo*, the nucleocapsid protein is usually present assembled into nucleocapsid structures composed of many hundred

associated N protein units surrounding the vRNA genome. The way in which the N protein interacts with P protein when it is assembled into nucleocapsid structures may well be different to the way in which it interacts when it is unassociated. In addition to the design of more highly resolved constructs, introduction of other nucleocapsid complex species into such an assay may lead to greater understanding of the interactions that occur in this elaborate association.

CHAPTER 11

CONCLUDING REMARKS.

11.1: INTRODUCTION.

Sequence homology comparisons suggest that the PVM P gene and P protein are closely related to the P genes and P proteins of the other pneumoviruses HRSV and TRTV, but show no homology to their counterparts in morbilliviruses or paramyxoviruses. Nucleotide sequence analysis of the PVM N gene has also revealed that, whilst the N gene and N proteins were highly homologous to their pneumovirus counterparts, they do not possess any significant sequence homology to the NP genes and NP proteins of morbilliviruses or paramyxoviruses (Barr *et al.*, 1991). Whilst these sequence homology comparisons show that the pneumovirus N and P proteins are different to their morbillivirus or paramyxovirus equivalents the pneumovirus nucleocapsid complex is believed to be organised and to function in a similar way to those of the morbilliviruses or paramyxoviruses. Whether this concept of the structure and function of the pneumovirus nucleocapsid by analogy with what is known of other negative strand viruses is warranted, is not yet clear as it is not supported by experimental data. This study investigated the expression characteristics of the N and P genes of PVM both *in vitro* and *in vivo* and attempted to characterise the polypeptides that these two genes are able to express. In addition, the interactions between the N and P proteins were investigated to characterise the nature of the association between the two major protein components of the nucleocapsid complex as an initial step in the characterisation of this structure which is central to the replication, transcription and assembly of the virus.

11.2 THE EXPRESSION CHARACTERISTICS OF THE PVM P GENE.

Work presented in this thesis has shown that *in vitro*, the PVM P gene long ORF is able to express the 39 kDa PVM P protein and at least four other polypeptides

with molecular weights of 37 kDa, 26 kDa, 23 kDa and 16 kDa. These four P protein related species were shown to be expressed by independent ribosomal initiation on internal AUG codons of the P gene mRNA. In addition to the 39 kDa P protein, the 26 kDa and 23 kDa P protein species identified *in vitro* also appeared to be expressed in PVM-infected BSC1 cells and this was supported by the results of anti-P protein monoclonal antibody epitope mapping studies. Formal confirmation that the *in vivo* expressed 26 kDa and 23 kDa polypeptides represented their *in vitro* expressed counterparts would require amino-terminal peptide sequence analysis. Seven other P protein related polypeptide species were also expressed in PVM-infected cells although the identity of these species is unknown. Because of their observed molecular weights it is unlikely that they represent primary expression products of internal initiation. Many of these species appear to be highly abundant and so peptide sequence analysis on these polypeptides to determine their identity may be possible. The 26 kDa and 23 kDa polypeptides are analogous to the X protein of SEN which is expressed from an internal AUG codon of the P protein ORF (Curran and Kolakofsky, 1988b). A polypeptide thought to represent the product of the PVM P gene second ORF (P2) was also identified in PVM-infected BSC1 cells. As above, formal confirmation that this polypeptide was expressed from the P gene second ORF would require peptide sequence analysis of this polypeptide species. The (P2) polypeptide is analogous to the C, C', Y1 and Y2 proteins of SEN. In common with the P genes of many other paramyxoviruses and morbilliviruses, it thus seems that the PVM P gene is able to direct the *in vivo* expression of multiple polypeptides from both the long and the short ORFs. The expression of all the polypeptides in addition to the PVM P protein appears to occur by way of independent ribosomal initiation on downstream AUG codons. In many paramyxoviruses and morbilliviruses the P gene coding region contained in the alternative V ORF is accessed by way of RNA editing which allows transcription of mRNAs possessing non-templated nucleotides. A sequence corresponding to the consensus paramyxovirus and morbillivirus P gene polymerase slippage site was not

identified on the PVM P gene suggesting that the second ORF is not accessed through the process of RNA editing. Similarly, analysis using anti-peptide antisera used to detect the second ORF protein suggested that frame-shifting between the two ORFs is unlikely to occur. However, given the dissimilarity between the pneumoviruses and the other members of the *Paramyxoviridae* family, the possibility that the PVM P gene uses a unique RNA editing mechanism to give rise to additional polypeptides should not be ruled out.

The presence of an extensive second ORF in the PVM P protein is a feature unique within the pneumovirus genus. It is not known what function the polypeptide product that is expressed from the P gene second ORF performs in the life cycle of PVM. The expression of a polypeptide from this ORF may be indicative of a fundamental difference that exists between processes that occur in the life cycle of PVM compared to the other pneumoviruses. The additional coding capacity of the PVM P gene over the P genes of other pneumoviruses may reflect a different requirement dictated by the PVM host cell environment. Alternatively, PVM may have evolved control mechanisms additional to those used by other pneumoviruses which are able to regulate processes such as virus assembly and RNA synthesis. If this is the case then the PVM P2 polypeptide would be functionally analogous to the SEN C and V polypeptides which have been shown to down regulate RNA synthesis. However it is possible that the P gene second ORF may supply PVM with a function that is expressed by the other pneumoviruses in a different viral component such as the L protein or the largely uncharacterised IC and IB polypeptides.

11.3: THE EXPRESSION CHARACTERISTICS OF THE PVM N GENE.

In vitro expression studies revealed that the PVM N gene was able to direct the synthesis of the PVM N protein and an additional and highly abundant 24 kDa polypeptide. This 24 kDa species was shown to be expressed by initiation on an internal N protein ORF AUG codon that is in an extremely favourable nucleotide context. *In vivo*, two products of the PVM N gene were also identified, the PVM N protein and a highly abundant 30 kDa polypeptide. The 24 kDa N-related polypeptide seen *in vitro* was not present *in vivo*. Because of its molecular weight, the 30 kDa polypeptide was thought not to represent a primary expression product of initiation on an internal AUG codon. More likely is that the 30 kDa species represents a modified expression product of internal initiation or alternatively a cleavage product, arising as a result of a specific protease cleavage event. Although unlikely, it is possible that it arises as a result of the PVM-infected BSC1 cell harvesting procedure. As with the other unidentified PVM proteins expressed *in vivo*, peptide sequence analysis could be used to reveal which part of the N protein the 30 kDa polypeptide represented, and pulse-chase labelling could be used to determine if the 30 kDa polypeptide arises by way of cleavage of full-length N protein.

11.4: POSSIBLE ROLES FOR THE ADDITIONAL N AND P GENE ENCODED PROTEINS.

The role played by the N, P and L proteins of *Paramyxoviridae* family members in the process of transcription and replication have been particularly well studied in SEN, and several functions and associations have been deduced for each nucleocapsid component. Amongst other associations there is much evidence to suggest that in VSV, the P protein binds to N protein to prevent N from assuming a self-aggregated form not capable of supporting genome replication (Howard and

Wertz, 1989). It is not known if the aggregated form of the SEN N protein performs any function. Therefore, any component that prevents the N and P proteins from binding to each other would affect genome replication. Whether the PVM N and P proteins are required to form a similar association that is essential for delivering the N protein in a replicational competent state is not certain although as this association is fundamental, it is likely that they do.

The additional amino-terminally truncated polypeptides expressed *in vivo* by the PVM P gene long ORF and also the 30 kDa N polypeptide may be able to affect PVM RNA synthesis by interfering with the association between full-length N and P proteins. With the use of a protein binding assay, it was determined that neither of the two *in vivo* expressed internally initiated P protein forms (26 kDa and 23 kDa) were capable of interacting with the PVM N protein. Therefore, these truncated P protein species are probably unable to influence the association of the N and P proteins and if they are effectors of RNA synthetic mechanisms they would have to act through association with a component other than the N protein, such as the L protein or the P protein itself.

Both the amino and carboxyl-terminal halves of the PVM N protein (amino acids 1-196 and 200-393 respectively) were found to be independently capable of binding P protein although with reduced binding affinities (5% and 20% respectively). This finding suggests that regardless of whether the 30 kDa N protein species is internally initiated or whether it is generated as a result of amino-terminal or carboxyl-terminal cleavage from full-length N protein, it will possess at least some PVM P protein binding ability. If the transcription and replication processes of PVM and SEN share a fundamental outline, presence of the 30 kDa N protein species in PVM-infected cells would result in a proportion of the free P protein being sequestered by the 30 kDa species and consequently being unavailable for binding to N protein. The quantity of N protein associated with P, and consequently the quantity of N protein available for

genome encapsidation would therefore decrease and the quantity of aggregated N protein would increase. According to current opinion, if this situation existed in the course of a SEN infection it would delay the switch between transcription and replication (Kolakofsky *et al.*, 1991). Expression of the 30 kDa species may also lead to an increase in the quantity of aggregated N protein, the function of which is uncertain. Since the P protein binding ability of the 30 kDa protein will be less than that for full-length N protein the postulated effect of the 30 kDa polypeptide may not be significant. These postulated functions may not apply to PVM because of the existence of uncharacterised proteins, such as the 1C and 1B proteins, which may be involved in mechanisms that control the polymerase complex in a wholly different way to that of SEN.

In conclusion, the work presented in this thesis has revealed that the PVM P gene possesses two extensive open reading frames and that both of these are able to express polypeptide products *in vivo*. The expression characteristics of the PVM N gene were also investigated and was found to direct the *in vivo* synthesis of the 43 kDa N protein and possibly a polypeptide with a molecular weight of 30 kDa. The ability of the PVM N and P proteins to associate with each other was studied and broad regions of these polypeptides that were involved in this interaction were identified.

CHAPTER 12
BIBLIOGRAPHY

- Anderson, K., Stott, E. J. and Wertz, G. W. (1992). Intracellular processing of the human respiratory syncytial virus fusion glycoprotein: Amino acid substitutions affecting folding, transport and cleavage. *Journal of General Virology* **73**, 1177-1188.
- Anderson, L. J., Heirholzer, J. C., Tsou, C., Hendry, R. M., Fernie, B. F., Stone, Y. and McIntosh, K. (1985). Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. *Journal of Infectious Diseases* **151**, 626-633.
- Banerjee, A. K. and Barik, S. (1992). Minireview: Gene expression of vesicular stomatitis virus genome RNA. *Virology* **188**, 417-428.
- Barik, S. and Banerjee, A. K. (1991). Cloning and expression of the vesicular stomatitis virus phosphoprotein gene in *Escherichia coli*: analysis of phosphorylation status versus transcriptional activity. *Journal of Virology* **65**, 1719-1726.
- Barik, S. and Banerjee, A. K. (1992). Sequential phosphorylation of the phosphoprotein of vesicular stomatitis virus by cellular and viral protein kinases is essential for transcription activation. *Journal of Virology* **66**, 1109-1118.
- Barr, J., Chambers, P., Pringle, C. R. and Easton, A. J. (1991). Sequence of the major nucleocapsid protein gene of pneumonia virus of mice; sequence comparisons suggest structural homology between nucleocapsid proteins of pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses. *Journal of General Virology* **72**, 677-685.

- Beckes, J. D. and Perrault, J. (1992). Stepwise phosphorylation of vesicular stomatitis virus P protein by virion associated kinases and uncoupling of second step from *in vitro* transcription. *Virology* **188**, 606-617.
- Bell, J. C. and Prevec, L. (1985). Phosphorylation sites on phosphoprotein NS of vesicular stomatitis virus. *Journal of Virology* **54**, 697-702.
- Berg, M., Hjertner, B., Moreno-Lopez, J. and Linne, T. (1992). The P gene of the porcine paramyxovirus LPMV encodes three possible polypeptides P, V and C: the P protein is edited. *Journal of General Virology* **73**, 1195-1200.
- Berthiaume, L., Joncas, J. and Pavilanis, V. (1974). Comparative structure, morphogenesis and biological characteristics of the respiratory syncytial (RS) virus and the pneumonia virus of mice (PVM). *Archive für die gesamte virusforschung* **45**, 39-51.
- Black, B. L. and Lyles, D. S. (1992). Vesicular stomatitis virus matrix protein inhibits host cell-directed transcription of target genes *in vivo*. *Journal of Virology* **66**, 4058-4064.
- Black, B. L., Rhodes, R. B., McKenzie, M. and Lyles, D. S. (1993). The role of vesicular stomatitis virus matrix protein in inhibition of host-directed gene expression is genetically separable from its function in virus assembly. *Journal of Virology* **67**, 4814-4821.
- Blixenkrone, Moller, M., Sharma, B., Varsanyi, T. M., Hu, A., Norrby, E. and Kovamees, J. (1992). Sequence analysis of the genes encoding the nucleocapsid

protein and phosphoprotein (P) of phocid distemper virus, and editing of the P gene transcript. *Journal of General Virology* **73**, 885-893.

Blumberg, B. M., Leppert, M. and Kolakofsky, D. (1981). Interaction of VSV leader RNA and nucleocapsid protein may control VSV genome replication. *Cell* **23**, 837-845.

Blumberg, B. M., Giorgi, C. and Kolakofsky, D. (1983). N protein of vesicular stomatitis virus selectively encapsidates leader RNA *in vitro*. *Cell* **32**, 559-567.

Blumberg, B. M., Chan, J. and Udem, S. A. (1991) Function of 3' and 5' end sequences: In theory and practice in *The Paramyxoviruses*, ed. Kingsbury, D. W. (Plenum Press, New York and London), pp. 235-247.

Boeck, R., Curran, J., Matsuoka, Y., Compans, R. and Kolakofsky, D. (1992). The parainfluenza virus type 1 P/C gene uses a very efficient GUG codon to start its C' protein. *Journal of Virology* **66**, 1765-1768.

Bucher, D. J., Kharitonov, I. G., Zakomirdin, J. A., Grigoriev, V. B., Klimenko, S. M. and Davis, J. F. (1980). Incorporation of influenza virus M protein into liposomes. *Journal of Virology* **36**, 586-590.

Buckland, R. and Wild, F. (1989). Leucine zipper motif extends. *Nature* **338**, 547.

Callain, P. and Roux, L. (1993). The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *Journal of Virology* **67**, 4822-4830.

- Cane, P. A., Matthews, D. A. and Pringle, C. R. (1991). Identification of the variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. *Journal of General Virology* **72**, 2091-2096.
- Canter, D. M., Jackson, R. L. and Perrault, J. (1993). Faithful and efficient *in vitro* reconstitution of vesicular stomatitis transcription using plasmid encoded L and P proteins. *Virology* **194**, 518-529.
- Caravokyri, C. and Pringle, C. R. (1992). Effect of changes in the nucleotide sequence of the P gene of respiratory syncytial virus on the electrophoretic mobility of the P protein. *Virus Genes* **6**, 53-62.
- Caravokyri, C., Zajac, A. J. and Pringle, C. R. (1992). Assignment of mutant tsN19 (complementation group E) of respiratory syncytial virus to the P protein gene. *Journal of General Virology* **73**, 865-873.
- Carthew, P. and Sparrow, S. (1980). Persistence of pneumonia virus of mice and Sendai virus in germ-free (nu/nu) mice. *British Journal Experimental Pathology* **61**, 172-175.
- Carthew, P., Sparrow, S. and Verstraete, A. P. (1978). Incidence of natural virus infections of laboratory animals 1976-1977. *Laboratory Animals* **12**, 245-246.
- Cash, P., Wunner, W. H. and Pringle, C. R. (1977). A comparison of the polypeptides of human and bovine respiratory syncytial virus and murine pneumonia virus. *Virology* **82**, 369-379.

- Cash, P., Preston, C. M. and Pringle, C. R. (1979). Characterization of murine pneumonia virus proteins. *Virology* **96**, 442-452.
- Castadenda, S. J. and Wong, T. C. (1989). Measles virus synthesises both leaderless and leader containing polyadenylated RNAs *in vitro*. *Journal of Virology* **63**, 2977-2986.
- Cattaneo, R., Schmid, A., Rebmann, G. and Baczko, K. (1986). Accumulated measles virus mutations in a case of subacute sclerosing panencephalitis : interrupted matrix protein reading frame and transcription alteration. *Virology* **154**, 97-107.
- Cattaneo, R., Kaelin, K., Baczko, K. and Billeter, M. A. (1989). Measles virus editing provides an additional cysteine-rich protein. *Cell* **56**, 759-764.
- Chambers, P., Barr, J., Pringle, C. R. and Easton, A. J. (1990a). Molecular cloning of pneumonia virus of mice. *Journal of Virology* **64**, 1869-1872.
- Chambers, P., Matthews, D. A., Pringle, C. R. and Easton, A. J. (1990b). The nucleotide sequences of intergenic regions between nine genes of pneumonia virus of mice establish the physical order of these genes in the viral genome. *Virus Research* **18**, 263-270.
- Chambers, P., Pringle, C. R. and Easton, A. J. (1990c). Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *Journal of General Virology* **71**, 3075-3080.
- Chambers, P., Pringle, C. R. and Easton, A. J. (1991). Genes 1 and 2 of pneumonia virus of mice encode proteins which have little homology with the 1C and 1B

proteins of human respiratory syncytial virus. *Journal of General Virology* **72**, 2545-2550.

Chambers, P., Pringle, C. R. and Easton, A. J. (1992). Sequence analysis of the gene encoding the fusion glycoprotein of pneumonia virus of mice suggests possible conserved secondary structure elements in paramyxovirus fusion glyoproteins. *Journal of General Virology* **73**, 1717-1724.

Chan, J., Pifko-Hirst, S. and Uden, S. A. (1989) Where is the measles virus (+) strand leader RNA transcript? in *Genetics and pathogenicity of negative strand viruses*, eds. Mahy, B. and Kolakofsky, D. (Elsevier biomedical press, Amsterdam), pp. 222-231.

Chattopadhyay, D. and Banerjee, A. K. (1987). Phosphorylation within a specific domain of the phosphoprotein of vesicular stomatitis virus regulates transcription *in vitro*. *Cell* **49**, 407-414.

Chong, L. D. and Rose, J. K. (1993). Membrane association of functional vesicular stomatitis virus matrix protein *in vivo*. *Journal of Virology* **67**, 407-414.

Collins, P. L. (1991) The molecular biology of human respiratory syncytial virus (RSV) of the genus pneumovirus in *The Paramyxoviruses*, ed. Kingsbury, D. W. (Plenum Press, New York and London), pp. 103-162.

Collins, P. L. and Mottet, G. (1992). Oligomerization and post-translational processing of glycoprotein G of human respiratory syncytial virus: Altered O-glycosylation in the presence of brefeldin A. *Journal of General Virology* **73**, 849-863.

Collins, P. L. and Mottet, G. (1993). Membrane orientation and oligomerization of the small hydrophobic protein of human respiratory syncytial virus. *Journal of General Virology* **74**, 1445-1450.

Collins, P. L. and Wertz, G. W. (1985). The envelope associated 22K protein of human respiratory syncytial virus: nucleotide sequence of the mRNA and a related polytranscript. *Journal of Virology* **54**, 65-71.

Collins, P. L., Dickens, L. E., Buckler-White, A., Olmsted, R. A., Spriggs, M. K., Camargo, E. and van Wyke Coelingh, K. (1986). Nucleotide sequences from the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. *Proceedings of the National Academy of Sciences, USA* **83**, 4594-4598.

Collins, P. L., Olmsted, R. A., Spriggs, M. K., Johnson, P. R. and Buckler-White, A. J. (1987). Gene overlap and site-specific attenuation of transcription of the viral polymerase L gene of human respiratory syncytial virus. *Proceedings of the National Academy of Sciences, USA* **84**, 5134-5138.

Collins, P. L., Olmsted, R. A. and Johnson, P. R. (1990). The small hydrophobic protein of human respiratory syncytial virus: comparison between antigenic subgroups A and B. *Journal of General Virology* **71**, 1571-1576.

Collins, P. L., Mink, M. A. and Stec, D. S. (1991). Rescue of synthetic analogs of respiratory syncytial virus genomic RNA and effect of truncations and mutations on the expression of a foreign reporter gene. *Proceedings of the National Academy of Sciences, USA* **88**, 9663-9667.

Collins, P. L., Mink, M. A., Hill III, M. G., Camargo, E., Grosfeld, H. and Stec, D. S. (1993). Rescue of a 7502-nucleotide (49.3% of full-length) synthetic analog of respiratory syncytial virus genomic RNA. *Virology* **195**, 252-256.

Compans, R. W., Harter, D. H. and Choppin, P. W. (1967). Studies of pneumonia virus of mice (PVM) in cell culture. II. Structure and morphogenesis of the virus particle. *Journal of Experimental Medicine* **126**, 275-277.

Curran, J. and Kolakofsky, D. (1988a). Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. *EMBO Journal* **7**, 245-251.

Curran, J. and Kolakofsky, D. (1988b). Scanning independent ribosomal initiation of the Sendai virus X protein. *EMBO Journal* **7**, 2869-2874.

Curran, J. A. and Kolakofsky, D. (1991). Rescue of a Sendai virus DI genome by other parainfluenza viruses: implications for genome replication. *Virology* **182**, 168-176.

Curran, J., Boeck, R. and Kolakofsky, D. (1991a). The Sendai virus P gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. *EMBO Journal* **10**, 3079-3085.

Curran, J., De Melo, M., Moyer, S. and Kolakofsky, D. (1991b). Characterization of the Sendai virus V protein with an anti-peptide antiserum. *Virology* **184**, 108-116.

- Curran, J., Marq, J. B. and Kolakofsky, D. (1992). The Sendai virus nonstructural C proteins specifically inhibit viral mRNA synthesis. *Virology* **189**, 647-656.
- Curran, J., Homann, H., Buchholz, C., Rochat, S., Neubert, W. and Kolakofsky, D. (1993). The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation. *Journal of Virology* **67**, 4358-4364.
- Dayhoff, M. O. (1978) A model of evolutionary change in proteins. Matrices for detecting distant relationships. In Dayhoff, M. O. (Ed), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington D.C pp345-358.
- De, B. P. and Banerjee, A. K. (1985). Requirements and functions of vesicular stomatitis virus L and NS proteins in the transcription process *in vitro*. *Biochemical and Biophysical Research Communications* **126**, 40-49.
- De, B. P., Thornton, G. B., Luk, D. and Banerjee, A. K. (1982). Purified matrix protein of vesicular stomatitis virus blocks viral transcription *in vitro*. *Proceedings of the National Academy of Sciences, USA* **79**, 7137-7141.
- De, B. P., Lesoon, A. and Banerjee, A. K. (1991). Human parainfluenza virus type 3 transcription *in vitro*: role of cellular actin in mRNA synthesis. *Journal of Virology* **65**, 3268-3275.
- Dubovi, E. J. and Wagner, R. R. (1977). Spatial relationships of the proteins of vesicular stomatitis virus: induction of reversible oligomers by cleavable protein cross-linkers and oxidation. *Journal of Virology* **22**, 500-509.

- Eaton, M. D. and van Herick, W. (1944). Demonstration in cotton rats and rabbits of a latent virus related to pneumonia virus of mice. *Proceedings of the Society for Experimental Biology and Medicine* **57**, 89-92.
- Elango, N., Satake, M., Coligan, J. E., Norrby, E., Camargo, E. and Venkatesan, S. (1985). Respiratory syncytial virus fusion glycoprotein: nucleotide sequence of mRNA, identification of cleavage activation site and amino acid sequence of N-terminus of F1 subunit. *Nucleic Acids Research* **13**, 1559-1574.
- Emerson, S. U. and Schubert, M. (1978). Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proceedings of the National Academy of Sciences, USA* **84**, 5655-5659.
- Emerson, S. U. and Yu, Y.-H. (1975). Both NS and L proteins are required for *in vitro* RNA synthesis by vesicular stomatitis virus. *Journal of Virology* **15**, 1348-1356.
- Faaberg, K. S. and Peeples, M. E. (1988). Association of soluble matrix protein of Newcastle disease virus with liposomes is independent of ionic conditions. *Virology* **166**, 123-132.
- Fernie, B. F., Dapolito, G., Cote, P. J. and Gerin, J. L. (1985). Kinetics of synthesis of RS virus glycoproteins. *Journal of General Virology* **66**, 1983-1990.

- Francki, R. I. B., Fauquet, C. M., Knudson, D. L. and Brown, F. (1991) Classification and nomenclature of viruses. Fifth report of the international committee on taxonomy of viruses. Archives of Virology. Springer-Verlag, Wein, New York.
- Frankel, A. D., Brecht, D. S. and Pablo, C. O. (1988). Tat protein from human immunodeficiency virus forms a metal-linked dimer. *Science* **240**, 70-73.
- Galinski, M. S. and Wechsler, S. L. (1991) The molecular biology of the paramyxovirus genus in *The Paramyxoviruses*, ed. Kingsbury, D. W. (Plenum press, New York), pp. 41-82.
- Galinski, M. S., Mink, M. A., Lambert, D. M., Wechsler, S. L. and Pons, M. W. (1986). Molecular cloning and sequence analysis of the human parainfluenza 3 virus RNA encoding the nucleocapsid protein. *Virology* **149**, 139-151.
- Galinski, M. S., Troy, R. M. and Banerjee, A. K. (1992). RNA editing in the phosphoprotein gene of the human parainfluenza virus type 3. *Virology* **186**, 543-550.
- Gallione, C. J., Greene, J. R., Iverson, L. E. and Rose, J. K. (1981). Nucleotide sequences of the mRNAs encoding the vesicular stomatitis virus N and NS proteins. *Journal of Virology* **39**, 529-535.
- Gannon, J. and Carthew, P. (1980). Prevalence of indigenous viruses in laboratory animal colonies in the United Kingdom 1978-1979. *Laboratory Animals* **14**, 309-311.

- Garcia, J., Garcia-Barreno, B., Vivo, A. and Melero, J. A. (1993). Cytoplasmic inclusions of respiratory syncytial virus-infected cells: formation of inclusion bodies in transfected cells that coexpress the nucleoprotein, the phosphoprotein, and the 22K protein. *Virology* **195**, 243-247.
- Garcia-Barreno, B., Palomo, C., Penas, C., Delgado, C., Perez-Brena, P. and Melero, J. (1989). Marked differences in the antigenic structure of human respiratory syncytial virus F and G glycoproteins. *Journal of Virology* **63**, 925-932.
- Garcia-Barreno, B., Portela, A., Delgado, T. and Lopez, J. (1990). Frame shift mutations as a novel mechanism for the generation of neutralization resistant mutants of human respiratory syncytial virus. *EMBO Journal* **12**, 4181-4187.
- Garcia-Barreno, B., Delgado, T., Akerlind-Stopner, B., Norrby, E. and Melero, J. A. (1992). Location of the epitope recognized by monoclonal antibody 63G on the primary structure of human respiratory syncytial virus G glycoprotein and the ability of synthetic peptides containing this epitope to induce neutralizing antibodies. *Journal of General Virology* **73**, 2625-2630.
- Gimenez, H. B., Cash, P. and Melvin, W. T. (1984). Monoclonal antibodies to human respiratory syncytial virus and their use in comparison of different isolates. *Journal of General Virology* **65**, 963-971.
- Gruber, C. and Levine, S. (1985). Respiratory syncytial virus polypeptides. IV. The oligosaccharides of the glycoproteins. *Journal of General Virology* **66**, 417-432.

- Gupta, K. C. and Patwardhan, S. (1988). ACG, the initiator codon for a Sendai virus protein. *Journal of Biological Chemistry* **263**, 8553-8556.
- Haeuptle, M. T., Flint, N., Gough, N. M. and Dobberstein, B. (1989). A tripartite structure of the signals that determine protein insertion into the endoplasmic reticulum membrane. *Journal of Cell Biology* **108**, 1227-1236.
- Hamaguchi, M., Yoshida, T., Nishikawa, K., Naruse, H. and Nagai, Y. (1983). Transcriptive complex of Newcastle disease virus. I. Both L and P proteins are required to constitute an active complex. *Virology* **128**, 105-117.
- Hammond, D. C., Evans, R. K. and Lesnaw, J. A. (1992a). The L protein of VSV transcription complexes is specifically photolabelled by 5 azido uridine 5' triphosphate an analogue of the RNA polymerase substrate uridine 5' triphosphate. *Journal of General Virology* **73**, 61-66.
- Hammond, D. C., Haley, B. E. and Lesnaw, J. A. (1992b). Identification and charaterization of serine/threonine protein kinase activity intinsic to the L protein of vesicular stomatis virus New Jersey. *Journal of General Virology* **73**, 67-75.
- Harter, D. H. and Choppin, P. W. (1967). Studies on pneumonia virus of mice (PVM) in cell culture. I. Replication in baby hamster kidney cells and properties of the virus. *Journal of Experimental Medicine* **126**, 251-266.
- Hendricks, D. A., Baradaran, K., McIntosh, K. and Patterson, J. L. (1987). Appearance of a soluble form of the G protein of respiratory syncytial virus in fluids of infected cells. *Journal of General Virology* **68**, 1705-1714.

- Hendricks, D. A., McIntosh, K. and Patterson, J. L. (1988). Further characterization of the soluble form of the G glycoprotein of respiratory syncytial virus. *Journal of Virology* **62**, 2228-2233.
- Herman, R. C. (1986). Internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA yields a second protein. *Journal of Virology* **58**, 797-804.
- Herman, R. C. (1987). Characterization of the internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA. *Biochemistry* **26**, 8346-8350.
- Herman, R. C. (1989). Synthesis of respiratory syncytial virus RNA in cell-free extracts. *Journal of General Virology* **70**, 755-761.
- Higgins, D. G. and Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237-244.
- Hirano, A., Wang, A. H., Gombart, A. F. and Wong, T. C. (1992). The matrix proteins of neurovirulent subacute sclerosing panencephalitis virus and its acute measles virus progenitor are functionally different. *Proceedings of the National Academy of Sciences, USA* **89**, 8745-8749.
- Homann, H. E., Willenbrink, W., Buchholz, C. J. and Neubert, W. J. (1991). Sendai virus protein-protein interactions studied by a protein-blotting protein-overlay technique: Mapping of domains on NP protein required for binding to P protein. *Journal of Virology* **65**, 1304-1309.

- Hopp, T.J. and Woods, K. R (1981). Prediction of protein alignment determinants from amino acid sequences. *Proceedings of the National Academy of Sciences, USA* **78**, 3824-3828.
- Horikami, S. M. and Moyer, S. A. (1991). Synthesis of leader RNA and editing of the P mRNA during transcription by purified measles virus. *Journal of Virology* **65**, 5342-5347.
- Horsfall, F. L. and Curnen, E. C. (1946). Studies on pneumonia virus of mice (PVM). II. Immunological evidence of latent infection with the virus in numerous mammalian species. *Journal of Experimental Medicine* **83**, 43-64.
- Horsfall, F. L. and Ginsberg, H. S. (1951). The dependence of the pathological lesion upon the multiplication of pneumonia virus of mice (PVM). Kinetic relation between the degree of viral multiplication and the extent of pneumonia. *Journal of Experimental Medicine* **93**, 151-160.
- Horsfall, F. L. and Hahn, R. G. (1939). A pneumonia virus of Swiss mice. *Proceedings of the Society for Experimental Biology and Medicine* **40**, 684-686.
- Horsfall, F. L. and Hahn, R. G. (1940). A latent virus in normal mice capable of producing pneumonia in its natural host. *Journal of Experimental Medicine* **71**, 391-408.
- Horsfall, F. L., Curnen, E. C., Mirick, G. S., Thomas, L. and Ziegler Jr., J. E. (1943). A virus recovered from patients with primary atypical pneumonia. *Science* **97**, 289-291.

- Horvath, C. M. and Lamb, R. A. (1992). Studies on the fusion peptide of a paramyxovirus fusion glycoprotein: Roles of conserved residues in cell fusion. *Journal of Virology* **66**, 2443-2455.
- Howard, M. and Wertz, G. W. (1989). Vesicular stomatitis virus RNA replication: a role for the NS protein. *Journal of General Virology* **70**, 2683-2694.
- Hsu, C.-H. and Kingsbury, D. W. (1982). Topography of phosphate residues in Sendai virus proteins. *Virology* **120**, 225-234.
- Huang, Y. T. and Wertz, G. W. (1982). The genome of respiratory syncytial virus is a negative-stranded RNA that codes for at least seven mRNA species. *Journal of Virology* **43**, 150-157.
- Huang, Y. T., Collins, P. L. and Wertz, G. (1984) Identification of a new envelope-associated protein of human respiratory syncytial virus in *Nonsegmented negative strand viruses, Paramyxoviruses and Rhabdoviruses*, eds. Bishop, D. H. L. and Compans, R. W. (Academic Press, New York), pp. 365-368.
- Huang, Y. T., Romito, R. R., De, B. P. and Banerjee, A. K. (1993). Characterisation of the *in vitro* system for the synthesis of mRNA from human respiratory syncytial virus. *Virology* **193**, 862-867.
- Hunt, D. M., Smith, E. P. and Buckley, D. W. (1984). Aberrant polyadenylation by a vesicular stomatitis virus mutant is due to an altered L protein. *Journal of Virology* **52**, 515-521.

Johnson, P. R. and Collins, P. L. (1988). The fusion glycoproteins of human respiratory syncytial virus of subgroups A and B: sequence conservation provides a structural basis for antigenic relatedness. *Journal of General Virology* **69**, 2623-2628.

Johnson, P. R. and Collins, P. L. (1990). Sequence comparison of the phosphoprotein mRNAs of antigenic subgroups A and B of human respiratory syncytial virus identifies a highly divergent domain in the predicted protein. *Journal of General Virology* **71**, 481-485.

Johnson, P. R., Spriggs, M. K., Olmsted, R. A. and Collins, P. L. (1987). The G glycoprotein of human respiratory syncytial viruses of subgroups A and B : Extensive sequence divergence between antigenically related proteins. *Proceedings of the National Academy of Sciences, USA* **84**, 5625-5629.

Kaptur, P. E., Rhodes, R. B. and Lyles, D. S. (1991). Sequences of the vesicular stomatitis virus matrix protein involved in binding to nucleocapsids. *Journal of Virology* **65**, 1057-1065.

Kingsbury, D. W., Bratt, M. A., Choppin, P. W., Hanson, R. P., Hosaka, Y., Meulen, T., Norrby, E., Plowright, W., Rott, R. and Wunner, W. H. (1978). Paramyxoviridae. *Intervirology* **10**, 137-152.

Klug, A. and Rhodes, D. (1987). Zinc fingers: A novel protein motif for nucleic acid recognition. *Trends in Biological Sciences* **12**, 464-469.

- Kolakofsky, D., Vidal, S. and Curran, J. (1991) Paramyxovirus RNA synthesis and P gene expression in *The Paramyxoviruses*, ed. Kingsbury, D. W. (Plenum Press, New York and London) pp215-233.
- Kondo, K., Bando, H., Kawano, M., Tsurudome, M., Komada, H., Nishio, M. and Ito, Y. (1990). Sequencing analyses and comparison of parainfluenza virus type 4A and 4B NP protein genes. *Virology* **174**, 1-8.
- Kozak, M. (1986a). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292.
- Kozak, M. (1986b). Regulation of protein synthesis in virus infected animals. *Advances in Virus Research* **31**, 229-292.
- Kozak, M. (1986c). Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences, USA* **83**, 2850-2854.
- Kozak, M. (1987). An analysis of 5' non coding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* **15**, 8125-8148.
- Kozak, M. (1989). The scanning model for translation: an update. *Journal of Cell Biology* **108**, 229-241.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

- La Ferla, F. M. and Peluso, R. W. (1989). The 1:1 N-NS protein complex of vesicular stomatitis virus is essential for efficient genome replication. *Journal of Virology* **63**, 3852-3857.
- Lambden, P. R. (1985). Nucleotide sequence of the respiratory syncytial virus phosphoprotein gene. *Journal of General Virology* **66**, 1607-1612.
- Lambert, D. M. (1988). Role of oligosaccharides in the structure and function of respiratory syncytial virus glycoproteins. *Virology* **164**, 458-466.
- Leopardi, R., Hukkanen, V., Vainionpaa, R. and Salmi, A. A. (1993). Cell proteins bind to sites with the 3' non coding region and the positive strand leader sequence of measles virus RNA. *Journal of Virology* **67**, 785-790.
- Levine, S., Klaiber-Franco, R. and Paradiso, P. R. (1987). Demonstration that glycoprotein G is the attachment protein of respiratory syncytial virus. *Journal of General Virology* **68**, 2521-2524.
- Li, Y., Schubert, M., Wagner, R. R. and Kang, C. Y. (1993). Viral liposomes released from insect cells infected with recombinant baculovirus expressing the matrix protein of vesicular stomatitis virus. *Journal of Virology* **67**, 4415-4420.
- Lichtenstein, D. L., Wertz, G. W. and Ball, L. A. (1991) Mutational analysis of the signal/anchor domain of the glycoprotein G of human respiratory syncytial virus.
- Ling, R. (1988) Polypeptides of murine and avian pneumoviruses. Ph.D. thesis (University of Warwick).

- Ling, R. and Pringle, C. R. (1989a). Polypeptides of pneumonia virus of mice. I: Immunological cross-reactions and post-translational modifications. *Journal of General Virology* **70**, 1427-1440.
- Ling, R. and Pringle, C. R. (1989b). Polypeptides of pneumonia virus of mice. II: Characterization of the glycoproteins. *Journal of General Virology* **70**, 1441-1452.
- Ling, R., Easton, A. J. and Pringle, C. R. (1992). Sequence analysis of the 22K, SH and G genes of turkey rhinotracheitis virus and their intergenic regions reveals a gene order different from that of other pneumoviruses. *Journal of General Virology* **73**, 1709-1715.
- Lui, F. T., Zinneker, M., Hamaoka, T. and Katz, D. H. (1979). New procedures for preparation and isolation of conjugates of proteins and a synthetic copolymer of D-amino acids and immunochemical characterisation of such conjugates. *Biochemistry* **18**, 690-697.
- Luk, D., Sanchez, A. and Banerjee, A. K. (1986). Messenger RNA encoding the phosphoprotein (P) gene of human parainfluenza virus 3 is bicistronic. *Virology* **153**, 318-325.
- Lyles, D. S., Puddington, L. and McCreedy, B. J. (1988). Vesicular stomatitis virus M protein in the nuclei of infected cells. *Journal of Virology* **62**, 4387-4392.
- Lyn, D., Gill, D. S., Scroggs, R. A. and Portner, A. (1991). The nucleoproteins of human parainfluenza virus type 1 and Sendai virus share amino acid sequences

and antigenic and structural determinants. *Journal of General Virology* **72**, 983-987.

Mallipedi, S. and Samal, S. K. (1992). Sequence comparison between the phosphoprotein mRNAs of human and bovine respiratory syncytial virus identifies a divergent domain in the predicted protein. *Journal of General Virology* **73**, 2441-2444.

Marx, P. A., Portner, A. and Kingsbury, D. W. (1974). Sendai virion transcriptase complex: polypeptide composition and inhibition by virion envelope proteins. *Journal of Virology* **13**, 107-112.

Masters, P. S. and Banerjee, A. K. (1988). Resolution of multiple complexes of phosphoprotein NS with nucleocapsid protein N of vesicular stomatitis virus. *Journal of Virology* **62**, 2651-2657.

Matsuoka, Y., Curran, J., Pelet, T., Kolakofsky, D., Ray, R. and Compans, R. W. (1991). The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a cysteine-rich V protein. *Journal of Virology* **65**, 3406-3410.

McGinnes, L., McQuain, C. and Morrison, T. (1988). The P protein and the non-structural 38K and 29K proteins of Newcastle disease virus are derived from the same open reading frame. *Virology* **164**, 256-264.

Mellon, M. G. and Emerson, S. U. (1978). Rebinding of transcriptase component (L and NS protein) to the nucleocapsid template of vesicular stomatitis virus. *Journal of Virology* **27**, 560-567.

- Mercyk, N., Horikami, S. M. and Moyer, S. A. (1988). The vesicular stomatitis virus L protein possesses the mRNA methyl transferase activities. *Virology* **163**, 222-225.
- Mink, M. A., Stec, D. S. and Collins, P. L. (1991). Nucleotide sequences of the 3' leader and 5' trailer regions of human respiratory syncytial virus genomic RNA. *Virology* **185**, 615-624.
- Miyahara, K., Kitada, S., Yoshimoto, M., Matsumura, H., Kawano, M., Komada, H., Tsurudome, M., Kusagawa, S., Nishio, M. and Ito, Y. (1992). Molecular evolution of human paramyxovirus: Nucleotide sequence analyses of the human parainfluenza type 1 virus NP and M protein genes and construction of phylogenetic trees for all the human paramyxoviruses. *Archives of Virology* **124**, 255-268.
- Moyer, S. A. (1989). Replication of the genome RNAs of defective interfering particles of vesicular stomatitis and Sendai viruses using heterologous viral proteins. *Virology* **172**, 341-345.
- Moyer, S. A., Baker, S. C. and Lessard, J. L. (1986). Tubulin, a factor necessary for the synthesis of both Sendai virus and vesicular stomatitis virus RNAs. *Proceedings of the National Academy of Sciences, USA* **83**, 5405-5409.
- Moyer, S. A., Baker, S. C. and Horikami, S. M. (1990). Host cell proteins required for measles virus reproduction. *Journal of General Virology* **71**, 775-783.

- Moyer, S. A., Smallwood-Kentro, S., Haddad, A. and Prevec, L. (1991). Assembly and transcription of synthetic vesicular stomatitis virus nucleocapsids. *Journal of Virology* **65**, 2170-2178.
- Mufson, M. A., Orvell, C., Rafnar, B. and Norrby, E. (1985). Two distinct subtypes of respiratory syncytial virus. *Journal of General Virology* **66**, 2111-2124.
- Nagai, Y., Yoshida, T., Hamaguchi, M., Linuma, M., Maeno, K. and Matsumoto, T. (1978). Cross-linking of Newcastle disease virus (NDV) proteins. *Archives of Virology* **58**, 15-28.
- Navarro, J., Lopez-Otin, C. and Villanueva, N. (1991). Location of phosphorylated residues in human respiratory syncytial virus phosphoprotein. *Journal of General Virology* **72**, 1455-1459.
- Nörrby, E., Mufson, M. A. and Sheshberadaran, H. (1986). Structural differences between subtype A and B strains of respiratory syncytial virus. *Journal of General Virology* **67**, 2721-2729.
- Ohgimoto, S., Bando, H., Kawano, M., Okamoto, K., Kondo, K., Tsurudome, M., Nishio, M and Ito, Y. (1990). Sequence analysis of P gene of human parainfluenza type 2 virus: P and cysteine-rich proteins are translated by two mRNAs that differ by two nontemplated G residues. *Virology* **177**, 116-123.
- Olmsted, R. A. and Collins, P. L. (1989). The 1A protein of respiratory syncytial virus is an integral membrane protein present as multiple, structurally distinct species. *Journal of Virology* **63**, 2019-2029.

- Olmsted, R. A., Murphy, B. R., Lawrence, L. A., Elango, N., Moss, B. and Collins, P. L. (1989). Processing, surface expression, and immunogenicity of carboxy-terminally truncated mutants of G protein of human respiratory syncytial virus. *Journal of Virology* **63**, 411-420.
- Palomo, C., Garcia-Barreno, B., Penas, P. and Melero, J. A. (1991). The G protein of human respiratory syncytial virus: significance of carbohydrate side chains and the C-terminal end to its antigenicity. *Journal of General Virology* **72**, 669-675.
- Paterson, R. G. and Lamb, R. A. (1990). RNA editing by nucleotide insertion in mumps virus P gene mRNA transcripts. *Journal of Virology* **64**, 4137-4145.
- Pattnaik, A. and Wertz, G. W. (1990). Replication and amplification of defective interfering particle RNAs of vesicular stomatitis virus in cells expressing viral proteins from vectors containing cloned cDNAs. *Journal of Virology* **64**, 2948-2957.
- Pattnaik, A. and Wertz, G. W. (1991). Cells that express all five proteins of vesicular stomatitis virus from cloned cDNAs support replication, assembly, and budding of defective interfering particles. *Proceedings of the National Academy of Sciences, USA* **88**, 1379-1383.
- Patton, J. T., Davis, N. L. and Wertz, G. W. (1984). N protein alone satisfies the requirement for protein synthesis during RNA replication of VSV. *Journal of Virology* **49**, 303-309.
- Pearson, H. E. and Eaton, M. D. (1940). A viral pneumonia of Syrian hamsters. *Proceedings of the Society for Experimental Biology and Medicine* **45**, 677-679.

- Pelet, T., Curran, J. and Kolakofsky, D. (1991). The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. *EMBO Journal* **10**, 443-448.
- Peluso, R. W. and Moyer, S. A. (1988). Viral proteins required for the *in vitro* replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* **162**, 369-376.
- Poch, O., Sauvaget, I., Delarve, M. and Tordo, N. (1989). Identification of four conserved motifs among the RNA dependent polymerase encoding elements. *EMBO Journal* **8**, 3867-3874.
- Pringle, C. R. and Eglin, R. P. (1986). Murine pneumonia virus: Seroepidemiological evidence of widespread human infection. *Journal of General Virology* **67**, 975-982.
- Rima, B. K. (1989) Comparison of amino acid sequences of the major structural protein of the paramyxo- and morbilliviruses in *Genetics and pathogenicity of negative strand viruses*, eds. Kolakofsky, D. and Mahy, B. (Elsevier press, Amsterdam).
- Roberts, S. R., Lichtenstein, D., Ball, L. A. and Wertz, G. W. (1993) Secreted and anchored forms of the attachment protein of respiratory syncytial virus are synthesized from different translational initiation sites (on the same mRNA). Presented at the IXth International Congress of Virology, Glasgow, Scotland, U.K.

- Routledge, E. G., Willcocks, M. M., Morgan, L., Samson, A. C. R., Scott, R. and Toms, G. L. (1987). Expression of the respiratory syncytial virus 22K protein on the surface of infected HeLa cells. *Journal of General Virology* **68**, 1217-1222.
- Rueda, P., Delgado, T., Portella, A., Melero, J. A. and Garcia-Barreno, B. (1991). Premature stop codons in the G glycoprotein of human respiratory syncytial viruses resistant to neutralization by monoclonal antibodies. *Journal of Virology* **65**, 3374-3378.
- Ryan, K. W. and Kingsbury, D. W. (1988). Carboxyl-terminal region of Sendai virus P protein is required for binding to viral nucleocapsids. *Virology* **167**, 106-112.
- Ryan, K. W. and Portner, A. (1990). Separate domains of Sendai virus P protein are required for binding to viral nucleocapsid. *Virology* **174**, 515-521.
- Ryan, K. W., Murti, K. J. and Portner, A. (1990). Localisation of P protein binding sites on the Sendai virus nucleocapsid. *Journal of General Virology* **71**, 997-1000.
- Ryan, K. W., Morgan, E. M. and Portner, A. (1991). Two non contiguous regions of Sendai virus P protein combine to form a single nucleocapsid binding site. *Virology* **180**, 126-134.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-497.

- Samal, S. K., Zamora, M., McPhilips, T. H. and Mohanty, S. B. (1991). Molecular cloning and sequence analysis of bovine respiratory syncytial virus mRNA encoding the major nucleocapsid protein. *Virology* **180**, 453-456.
- Samal, S. K., Pastey, M. K., McPhilips, T. H. and Mohanty, S. B. (1993). Bovine respiratory syncytial virus nucleocapsid protein expressed in insect cells interacts with the phosphoprotein and the M2 protein. *Virology* **193**, 470-473.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning, a laboratory manual. Second edition. Cold Spring Harbour. Cold Spring Harbour press
- Samson, A. C. R., Leversley, I. and Russell, P. H. (1991). The 36 kDa polypeptide synthesised in Newcastle disease virus-infected cells possesses the properties predicted for the hypothesised 'V' protein. *Journal of General Virology* **72**, 1709-1713.
- Sanchez, A., De, B. P. and Banerjee, A. K. (1985). *In vitro* phosphorylation of NS protein by the L protein of vesicular stomatitis virus. *Journal of General Virology* **66**, 1025-1036.
- Sanderson, C. M., McQueen, N. L. and Nayak, D. P. (1993). Sendai virus assembly: M protein binds to viral glycoproteins in transit through the secretory pathway. *Journal of Virology* **67**, 651-663.
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161-178.

- Satake, M. and Venkatesan, S. (1984). Nucleotide sequence of the gene encoding the respiratory syncytial virus matrix protein. *Journal of Virology* **50**, 92-99.
- Satake, M., Elango, N. and Venkatesan, S. (1984). Sequence analysis of the respiratory syncytial virus phosphoprotein gene. *Journal of Virology* **52**, 991-994.
- Schwalbe, J. C. and Hightower, L. E. (1982). Maturation of the envelope glycoproteins of Newcastle disease virus on cellular membranes. *Journal of Virology* **41**, 947-957.
- Sherwood, S. and Moyer, S. A. (1993). Promoter analysis of the vesicular stomatitis virus RNA polymerase. *Virology* **192**, 254-263.
- Shioda, T., Iwasaki, K. and Shibuta, H. (1986). Determination of the complete nucleotide sequence of the Sendai virus genome RNA and the predicted amino acid sequences of the F, HN and L proteins. *Nucleic Acids Research* **14**, 1545-1563.
- Shipley, J. B., Pal, R. and Wagner, R. R. (1988). Antigenicity, function and conformation of synthetic oligopeptides corresponding to amino-terminal sequences of wild-type and mutant matrix proteins of vesicular stomatitis virus. *Journal of Virology* **62**, 2569-2577.
- Sleat, D. E., Chikkala, N. F., Gautman, S. and Banerjee, A. K. (1992). Restricted replication of vesicular stomatitis virus in T lymphocytes is coincident with a deficiency in a cellular protein kinase required for viral transcription. *Journal of General Virology* **73**, 3125-3132.

- Smith, A. L., Carrano, V. A. and Brownstein, D. G. (1984). Response of weanling random-bred mice to infection with pneumonia virus of mice (PVM). *Laboratory Animal Science* **34**, 35-37.
- Southern, J. A., Precious, B. and Randall, R. E. (1990). Two non-templated nucleotide additions are required to generate the P mRNA of parainfluenza virus type 2 since the RNA genome encodes protein V. *Virology* **177**, 388-390.
- Spriggs, M. K. and Collin, P. L. (1986). Sequence analysis of the P and C protein genes of human parainfluenza virus type 3 : Patterns of amino acid sequence homology among paramyxovirus proteins. *Journal of General Virology* **67**, 2705-2719.
- Stec, D. S., Hill III, M. G. and Collins, P. L. (1991). Sequence analysis of the polymerase L gene of human respiratory syncytial virus and predicted phylogeny of nonsegmented negative-strand viruses. *Virology* **183**, 273-287.
- Studier, F. L. and Moffat B. A. (1986). The use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. *Journal of Molecular Biology* **189**, 113-130.
- Sullender, W. M. and Wertz, G. W. (1991) The unusual attachment glycoprotein of the respiratory syncytial viruses: Structure, maturation, and role in immunity in *The Paramyxoviruses*, ed. Kingsbury, D. W. (Plenum Press, New York and London), pp. 383-406.

- Sullender, W. M., Anderson, K. and Wertz, G. W. (1990). The respiratory syncytial virus subgroup B attachment glycoprotein: Analysis of sequence, expression from a recombinant vector, and evaluation as an immunogen against homologous and heterologous subgroup virus challenge. *Virology* **178**, 195-203.
- Sullender, W. M., Mufson, M. A., Anderson, L. J. and Wertz, G. W. (1991). Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. *Journal of Virology* **65**, 5425-5434.
- Takacs, A. M., Perrine, K. G., Barik, S. and Banerjee, A. K. (1991). Alteration of specific amino acid residues in the acidic domain I of VSV phosphoprotein (P) converts a GAL4-P (I) hybrid into a transactivator. *New Biologist* **3**, 581-591.
- Takacs, A. M., Barik, S., Das, T. and Banerjee, A. K. (1992). Phosphorylation of specific serine residues within the acidic domain of the phosphoprotein of vesicular stomatitis virus regulates transcription *in vitro*. *Journal of Virology* **66**, 5842-5848.
- Takeuchi, K., Tanabayashi, K., Hishiyama, M., Yamada, Y. K., Yamada, A. and Sugiura, A. (1990). Detection and characterization of mumps virus V protein. *Virology* **178**, 247-253.
- Thomas, S. M., Lamb, R. A. and Paterson, R. G. (1988). Two mRNAs that differ by two non-templated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell* **54**, 891-902.

- Thuring, R. W. J., Sanders, J. P. M. and Borst, P. (1975). A freeze-squeeze method for recovering DNA from agarose gels. *Analytical Biochemistry* **66**, 213-220.
- Vidal, S. and Kolakofsky, D. (1989). Modified model for the switch from Sendai virus transcription to replication. *Journal of Virology* **63**, 1951-1958.
- Vidal, S., Curran, J. and Kolakofsky, D. (1990a). Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus-encoded activity. *Journal of Virology* **64**, 239-246.
- Vidal, S., Curran, J. and Kolakofsky, D. (1990b). A stuttering model for paramyxovirus P mRNA editing. *EMBO Journal* **9**, 2017-2022.
- Walravens, K., Kettermann, R., Collard, A., Coppe, D. and Burney, A. (1990). Sequence comparisons between the fusion protein of human and bovine respiratory syncytial viruses. *Journal of General Virology* **71**, 3009-3014.
- Walsh, E. E., Brandriss, M. W. and Schlesinger, J. J. (1987). Immunological differences between the envelope glycoproteins of two strains of human respiratory syncytial virus. *Journal of General Virology* **68**, 2169-2176.
- Weir, E. C., Brownstein, D. G., Smith, A. L. and Johnson, E. A. (1988). Respiratory disease and wasting in athymic mice infected with pneumonia virus of mice. *Laboratory Animal Science* **38**, 133-137.
- Wertz, G. W. and Levine, M. (1973). RNA synthesis by vesicular stomatitis virus and a small plaque mutant: effects of cycloheximide. *Journal of Virology* **12**, 253-264.

- Wertz, G. W., Krieger, M. and Ball, L. A. (1989). Structure and cell surface maturation of the attachment glycoprotein of human respiratory syncytial virus in a cell line deficient in O-glycosylation. *Journal of Virology* **63**, 4767-4776.
- Yamada, H., Hayata, S., Omata-Yamada, T., Taira, H., Mizumoto, K. and Iwasaki, K. (1990). Association of Sendai virus C protein with nucleocapsids. *Archives of Virology* **113**, 245-253.
- Yu, Q., Davis, P. J., Barrett, T., Binns, M. M., Boursnell, M. E. G. and Cavanagh, D. (1992a). Deduced amino acid sequence of the fusion glycoprotein gene of turkey rhinotracheitis virus has greater identity with that of human respiratory syncytial virus, a pneumovirus, than that of paramyxoviruses or morbilliviruses. *Journal of General Virology* **72**, 75-81.
- Yu, Q., Davis, P. J., Brown, T. D. K. and Cavanagh, D. (1992b). Sequence and *in vitro* expression of the M2 gene of turkey rhinotracheitis pneumovirus. *Journal of General Virology* **73**, 1355-1363.
- Yu, Q., Davis, P. J., Li, J. and Cavanagh, D. (1992c). Cloning and sequencing of the matrix protein (M) gene of turkey rhinotracheitis virus reveal a gene order different from that of respiratory syncytial virus. *Virology* **186**, 426-434.
- Zakowski, J. J., Petri, W. A. and Wagner, R. R. (1981). Role of the matrix protein in assembly the membrane of vesicular stomatitis virus: Reconstitution of matrix protein with negatively charged phospholipid vesicles. *Biochemistry* **20**, 3902-3907.

Zamora, M. and Samal, S. K. (1992). Sequence analysis of M2 mRNA of bovine respiratory syncytial virus obtained from an F-M2 dicistronic mRNA suggests structural homology with that of human respiratory syncytial virus. *Journal of General Virology* **73**, 737-741.

Zuker, M (1989). On finding all suboptimal foldings of an mRNA molecule. *Science* **244**, 48-52.